University of Reading

Exploring the Relationship between Vocabulary Knowledge and Reading Comprehension in L1 Arabic Learners of English

UNIVERSITY OF READING

PhD Thesis

Reham Alkhudiry

Prof. Jeanine Treffers-Daller

Dr. Jacqueline Laws

January 2018

Table of Content

List of Figures

List of Tables

List of Abbreviations

AWL	Academic Word List
BNC	British National Corpus
CPE	Cambridge-ESOL Proficiency Certificate in English
CEFR	Common European Framework of Reference
DST	Digit Span Test
DRP	Degrees of Reading Power
DPT	Delay Prime Test
DVK	Depth of Vocabulary Knowledge
EFL	English as a Foreign Language
ESL	English as a Second Language
ET	Elaboration Task
G1	Group 1
G2	Group 2
GO	Graphic Organizer
GSL	General Service List
HF	High Frequency
IELTS	International English Language Testing System
IH	Input Hypothesis
ILH	Involvement Load Hypothesis
IPT	Immediately Prime Test
IQ	Intelligence Quotient
L1	First Language
L2	Second Language
LDT	Lexical Decision Task
LF	Low Frequency
LTM	Long-term memory
MC	Multiple-Choice
MSA	Modern Standard Arabic
NS	Native Speaker
RCQ	Reading Comprehension Question
RSQ	Reading Summarization Question
RT	Reaction Time

SD	Standard Deviation
SLA	Second Language Acquisition
SOA	Stimulus Onset Asynchrony
STM	Short-term memory
SPT	Semantic Prime Test
SVR	Simple View of Reading
SWRT	Single Word Reading Test
TOFEL	Test of English as Foreign Language
VKS	Vocabulary Knowledge Scale
VLT	Vocabulary Level Test
VST	Vocabulary Size Test
WAT	Word-Associate Test
WMS	Working Memory Size
WAIS	Wechsler-Adult Intelligence Scale
WR	Written Recall
YARK	York assessment Reading Comprehension

List of Appendix

Acknowledgments

First and foremost, thanks should be expressed to my God (Allah), for giving me the opportunity of being engaged in advanced studies in the UK. I would like to express my appreciation to Qassim University, the Ministry of Higher Education in Saudi Arabia, and the Saudi government, particularly King Salman bin Abdul-Aziz Al-Saud for the scholarship granted to enable this research to be undertaken.

I would also like to express my most sincere thanks to my first supervisor Prof. Jeanine Treffers-Daller for her valuable supervision, inspiring guidance in the area of vocabulary acquisition, and for constant encouragement. My great thanks should also be expressed to my second supervisor Dr. Jacqueline Laws for her appreciated assistance in using the SuperLab and the SPSS softwares, and constructive feedback on my thesis chapters. I have been very fortunate to experience the inspiration of my excellent supervisors whom I have learnt many key skills related to my research and writing this thesis. My great thanks further are due to the academic and administrative staff in the English Language and Applied Linguistics department at the University of Reading whose warm welcome and supportive directions have made my PhD study easier. Moreover, I am deeply grateful to the university students who gave me part of their busy timetable to volunteer and take part in both the piloting procedures and this experimental research. I thank them for their patience and co-operation.

Last but not least, my deepest appreciation must go to my husband and my daughters for their care and support throughout my academic pursuit. They shared the highs and lows of this seemingly epic journey with me and hardly ever complained during my absence. Special thanks go to my mother for her constant emotional support and encouragement. I would also like to thank my sisters, brothers, and friends for their support throughout the period of my doctoral research. This work is dedicated to my mother and my husband, whose support and company made this possible, and also to my father who supported me throughout my study journey, unfortunately, is no longer with me to celebrate the completion of my PhD thesis.

Abstract

Developing a large enough vocabulary is an essential element in L2 acquisition in order to be able to read and write and participate in various topics of conversation. It is also assumed that reading is an important skill for academic success in first language (L1) and second language (L2) learning (Elley, 1991; Pulido, 2003). Many studies have focused on the contribution of vocabulary size to reading in L1 and L2 (e.g., Nation, 2006; Hsueh-chao & Nation, 2000), and conversely, on the extent to which learners can learn words incidentally from reading (Horst, Meara & Cobb, 1998; Waring & Takaki, 2003; Pigada & Schmitt, 2006). However, we know much less about the relationship between the depth of vocabulary knowledge and reading comprehension (Qian, 2002). Jiang (2000) proposed a three-stage model of lexical development in L2: at the first stage, the L2 learner's focus is on the formal aspects of the word, then at the second stage the L1 meaning and syntactic information is transferred into the L2 lexeme, and at the last stage, the L2 semantic syntactic information can be integrated into the new L2 lexical entry. Study 1 focuses on a) exploring which aspects of words (meaning, form and use) L2 learners can retain through reading, based on Jaing's (2000) model, b) how depth and size of vocabulary knowledge can explain variance in reading comprehension performance, and c) whether frequency of occurrence explains learning and retention of new words from reading. Two offline tests were used to measure depth and size of word knowledge among 30 L1 Arabic learners of English and 30 native English speakers, and one online lexical decision task measured form recognition. The target words were four non-words which replaced four existing nouns in two stories from the YARC reading comprehension test. The results show that the L2 learners are better at retaining word forms than word meanings of these non-words. Vocabulary depth (knowledge about a word's meaning and use) explains a significant variance in summarising the stories from the YARC. The target non-words occurring more frequently (eight times) have positively explained the form recognition of these words, however, this appears to decline over time for the L1 Arabic learners of English group.

Study 2 is an intervention study, designed after obtaining Study 1 results in order to examine the role of a specific technique in explaining L2 retention of the meaning of new words among 40 L1 Arabic learners of English. It is based on Laufer and Hulstijn's (2001) Involvement Load Hypothesis (ILH) and examines how an elaboration task (ET), as a post reading activity with high involvement component, can improve L2 readers' retention of the meaning of new words. Based on Jiang's (2000) model, it further aims to investigate connections between L2 words and conceptual representation, and particularly whether the

group which had carried out an elaboration task with these target non-words were better at gaining the L2 lemma than those who had not carried out this task. It also investigates whether frequency of occurrence explained students' ability to learn words from reading. The same two stories from the YARC were used, in which the same four target non-words replaced existing words. An offline test was used to measure knowledge of meaning and use of the target non-words and one online Semantic Priming Task measured respondents' accuracy and speed in recognising the meaning of target non-words. The regression analysis shows that an ET with a high involvement component significantly contributes to L2 learning and retention of the meaning of new words. Based on Jiang's (2000) model, L2 learners could link the L2 meaning to the L2 English form, but this seems to disappear one week later. In terms of the ET groups, L2 who took the ET involving the target non-word continued to link the L2 lemma semantic information to the target L2 form, but this also appears to decline over time. Target non-words occurring more frequently are better retained, however, this disappears one week later. These findings have significant theoretical and pedagogical implications for enhancing L2 vocabulary learning and retention.

Chapter 1: Introduction

1.1 Overview

This chapter provides an overview of the present study. It describes the background of and rationale for the current study and states its significance and general aims. The chapter also discusses L2 learning difficulties, particularly in the case of Arabic learners of English. Lastly, it outlines the structure of the thesis.

1.2 Background of the Study

Reading has always been regarded as an important skill for academic success in first-language (L1) and second-language (L2) learning (Elley, 1991; Johns, 1981; Pulido, 2003). Many variables influence reading comprehension in both L1 and L2, such as background knowledge, word knowledge of the text, and application of reading strategies, including recognising the sentence pattern, comprehending the main idea, and making inferences (Laufer, 1997b). Krashen's Input Hypothesis (IH) (1985) advises that written language input should be comprehensible in order to allow the underlying linguistic competence to develop. Krashen (1985) defines comprehensible input as the learners' target language that cannot be produced, but can still be understood. In other words, the input is considered to be comprehensible when the learner is able to understand the message (what is written) rather than its form (how is written) (Park, 2005). What makes an input comprehensible? There has been significant research, developed from Krashen's (1985) Input Hypothesis, that identifies the features that impact the comprehensibility of a text while reading, such as the following:

Prior vocabulary knowledge: Krashen (1989) notes that learners use their existing linguistic skill, together with their word knowledge, to deduce the meaning of the input in order to find the text more comprehensible. This means that L2 learners need to acquire a substantial L2 vocabulary in order to be able to understand L2 written material. Without a sufficiently large vocabulary, the ability to guess the meaning of written words through context can be difficult for many learners.

Familiarity: This is an influential concept in facilitating the comprehensibility of the written input. Preview questions, activities, pictures, or visual aids that introduce the topic and render the texts more familiar to L2 learners can aid their comprehension. Similarly, a series of related topics can enhance learners' familiarity and then encourage them to understand the written input (Scarcella, 1997).

Using material appropriate to learners' English level: Krashen (1985) indicates using written input appropriate to the English level of the learners can enhance their overall English language development. As suggested, written input, that is neither too difficult nor too easy for the learner, can facilitate English language acquisition. Learners would be able to understand more information when the text is adapted according to their English levels.

Culture-specific knowledge: L2 learners may often find difficulties if the written input is drawn from a different cultural source or context. In such cases, they may lack adequate background knowledge of the topic, which may make the input less comprehensible. Thus, English learners should draw on experiences consistent with their cultures to make the input easier to understand.

Narrow reading: Reading a large number of materials on a specific topic can help L2 learners improve their English level because reading narrowly seems to increase the chances of L2 learners encountering similar vocabulary and grammatical structures, and repeated exposure to those words and structures may facilitate their English language acquisition. Other features such as *reduced length, easy reading for pleasure* with the aid of *pictures* can help make the input more comprehensible to beginning learners of English (Scarcella, 1997).

Krashen (1989) further argued that reading leads to incidental vocabulary acquisition, that is, whenever reading provides comprehensible and sufficient input, it eventually leads to acquisition. Incidental learning is an important factor in enhancing learning vocabulary incidentally from context, as it can be argued that L2 learners may not be able to acquire an adequate level of new vocabulary via explicit teaching only, due to the fact that classroom periods for vocabulary instruction are short. Incidental learning refers to learning which takes place without specific intention, such as learning new vocabulary while engaging in any language tasks or activities (Nation, 2001). One of the early studies of vocabulary learning in L1 (Nagy, Herman, & Anderson, 1985), focused on the importance of learning new vocabulary from context. It was found that a single incidental exposure of a word may lead to a small gain in vocabulary knowledge. Moreover, this process of incidental learning of words seems to be a gradual process, as claimed by Anderson (1985; cited in Richards & Renandya, 2002). Therefore, Nagy et al. (1985) suggested that when this learning process is enhanced by exposure to a sufficient amount of written input, incidental word learning can be significantly increased. Furthermore, findings from L2 studies suggested that L2 vocabulary can also be gained incidentally through repeated exposure to the target words (Huckin & Coady, 1999; Gass, 1999; Waring & Takaki, 2003; Horst, 2005).

Reading, therefore, has been considered an important and effective element for encouraging incidental learning of L2 vocabulary. Krashen (1989) specifies that vocabulary and spelling acquisition can be achieved through exposure to comprehensible input via incidental reading. Thus, many scholars have stressed the usefulness of learning new words using comprehensible written input, and this notion has become widely accepted in L2 research (Dupuy & Krashen, 1993; Horst, 2005; Pigada & Schmitt, 2006). Hunt and Beglar (1998) further clarified that learning new words incidentally can be a useful and effective strategy for all language learners at various levels. They pointed out that encouraging learners to read can provide them with opportunities to learn new words. Huckin and Coady (1999) summarised some advantages of incidental vocabulary learning from reading. It is a) a context learning in which a learner encounters a rich source of vocabulary use and meaning, b) practically effective since it facilitates vocabulary learning at the same time as reading, c) more learner-based, as learners have their own choice in selecting their reading materials.

Vocabulary acquisition is considered by many scholars to be just as important as other language skills (Coady & Huckin, 1997), as well as being a key element in L2 language acquisition (Laufer, 1992). Thus, building a large vocabulary helps L2 learners to read, write, and communicate with others on various topics of conversation. One of the major problems that L2 learners of English may face, however, is how to go about learning the L2 words they need (Laufer, 1992). Several studies have investigated the number of words that native English speakers and L2 learners of English need to know in order to be able to communicate effectively (Goulden, Nation, & Read, 1990; Schmitt, 2010) and to read adequately (Laufer & Sim, 1985; Laufer, 1989; Zechmeister, Chronis, Cull, D'Anna, & Healy, 1995; Hsueh-chao & Nation, 2000; Schmitt, Xiangying, & William, 2011). Findings in second language acquisition (SLA) research have shown that the average five-year-old child begins school with a vocabulary of between 4,000 to 5,000 word families; and graduates from university with a vocabulary of about 20,000 word families. Schmitt (2010) confirms this by estimating that a range of 16,000 - 20,000 word families is typical for educated L1 English native speakers.

In terms of L2 learners' vocabulary size, Hsueh-chao and Nation (2000) and Nation (2006) conclude that 8,000 - 9,000 word families may be adequate for L2 learners to be able to comprehend 98% of English texts. Nation (2006) also estimates that L2 learners need to know 6000 - 7000 word families in order to be able to deal with daily conversation in English. Laufer (1997a; 2000), on the other hand, suggests that a lower target of 3,000 - 5,000 word families can be sufficient for reading by L2 learners. It seems that both native speakers and L2 learners

of English vary in their vocabulary size, depending on how they use their language and how much.

Read (2000) noted that vocabulary knowledge is an essential requirement in the language learning process and clarified that "vocabulary is not just a set of linguistic units but also an attribute of individual language learners, in the form of vocabulary knowledge and the ability to access that knowledge for communicative purposes" (p. 14). Therefore, both teachers and learners should be aware of the importance of developing an L2 lexicon in L2 learning. Jiang (2000) proposed a three-stage model of lexical development in a second language. At the first stage, the L2 learner's focus is on the formal aspects of the word, including spelling and pronunciation. At the second stage, because L2 learners have already established L1 lexical and semantic systems, the L1 meaning and syntactic information is transferred into the L2 lexeme. At the last stage, the L2 semantic, syntactic, and morphological information is integrated into the new L2 lexical entry. Jiang also claimed that most of the L2 words become fossilised in the second stage (the L1 lemma mediation stage). A detailed discussion of this model and how L2 lexical items are organised and processed in the mental lexicon of a learner will be presented in Chapter 2 (Section 2.4.2.1).

With reference to the importance of vocabulary knowledge as an effective element in reading comprehension, research has paid more attention to size of vocabulary (e.g., Liu & Nation, 1985; Laufer, 1992) than depth of vocabulary knowledge (e.g., Paribakht & Wesche, 1997; Qian, 1999; 2002). Rashidi and Khosravi (2010) have argued that depth of vocabulary knowledge contributes more to reading comprehension than vocabulary size. Others have placed, however, greater emphasis on the size of vocabulary in order to achieve success in reading comprehension (Liu & Nation, 1985; Laufer, 1992; Hsueh-chao & Nation, 2000). Relative contributions of size and depth to reading comprehension, therefore, remain unclear, in part due to the way in which these two dimensions are operationalised.

Although it is generally agreed that incidental vocabulary learning via reading is possible, L2 learners need more support to enhance incidental vocabulary learning. Researchers, therefore, have offered different approaches for enhancing incidental vocabulary learning from reading. As an example, Laufer and Hulstijn (2001) proposed the Involvement Load Hypothesis (ILH), which builds on several earlier studies investigating the effectiveness of different tasks in enhancing word learning (Hulstijn, 1992; Cho & Krashen, 1994; Hulstijn & Trompetter, 1998). Laufer and Hulstijn (2001) claimed that greater depth of processing will lead to better retention of new words. In other words, learning and retention of L2 words is influenced by the

involvement load induced by a task, as determined by three components: need, search, and evaluation. A deeper level of processing new words can be obtained by drawing more attention to different aspects of the meaning, form and use of a word and its linkages to other words. Also, processing these words more fully will lead to higher retention than by processing new information with less involvement.

A number of researchers have examined the validity of the Involvement Load Hypothesis (ILH) proposed by Laufer and Hulstijn (2001) and have provided empirical evidence supporting this theory. They confirmed that a vocabulary task which induced a higher involvement load leads to better retention (Keating, 2008; Kim, 2008; 2011; Eckerth & Tavakoli, 2012; Tang & Treffers-Daller, 2016). On the other hand, researchers such as Folse (2006) and Yaqubi, Rayati and Allemzade Gorgi (2010) have challenged the ILH and argued that a lower involvement index leads to better results than a higher index. Folse (2006) claimed that word learning is influenced more by repeated exposure than by depth of word processing. However, this research is limited compared to the broad set of investigations that support this theory. This, indeed, confirms Laufer and Hulstijn's (2001) argument that further research to support the claim made by ILH is greatly needed in order to determine the most effective approaches for enhancing incidental vocabulary learning through reading.

In summary, teaching and learning L2 vocabulary have been widely studied from different perspectives in L2 acquisition research (Knight, 1994; Hulstijn, Hollander, & Greidanus, 1996; Hunt & Beglar, 1998). Nevertheless, the relevance of reading has attracted the most attention from researchers as an effective approach for acquiring L2 vocabulary incidentally (Horst, Cobb, & Meara, 1998; Waring & Takaki, 2003; Horst, 2005; Pigada & Schmitt, 2006; Brown, Waring, & Donkaewbua, 2008). Therefore, it has become important to investigate the effect of this approach on L2 learners of English in order to develop learners' reading skills and to enrich their vocabulary.

1.3 The Study: Rationale and Aims

The current research is basically driven by two rationales: the researcher's professional interest and the gap in the literature of L2 vocabulary acquisition (see Chapter 3, Section 3.1 and Chapter 4, Section 4.1). As a lecturer in the English Language and Applied Linguistics Department at Qassim University in Saudi Arabia, I noticed that students struggled with reading because they encounter so many unknown words. This encouraged me to follow this field through my PhD study and focus on exploring ways to effectively increase and expand vocabulary knowledge of L1 Arabic learners of English. The present study is also conducted

with the aim of filling the gaps in L2 vocabulary acquisition literature by linking some previously unconnected theories and by answering some remaining questions in the literature, as will be shown in the next paragraph.

The current study, therefore, is divided into two parts: Study 1 investigates the relationship between vocabulary knowledge and reading comprehension performance, whereas Study 2 is an intervention study aimed at examining the role of a post reading task with a high involvement load in enhancing L2 learning and retention of the meaning of new words encountered through reading.

By conducting an experimental study at Qassim University with L1 Arabic learners of English, Study 1 explored the ways in which reading could contribute to L2 vocabulary learning and retention and the ways in which depth and size could contribute to reading comprehension performance. Study 1 is based on Jiang's (2000) model, and, to the best of the researcher's knowledge, Jiang's model has, so far, not been applied to studies of the links between vocabulary depth and reading comprehension. In this sense, Study 1 aims to explore those aspects of words which L2 learners can acquire and retain through reading comprehension, based on meaning, form, and use of words in a sentence. As the relative contribution of size and depth to reading comprehension remains ambiguous (Qian, 1999; 2002), an additional aim of Study 1 was to identify the extent to which the learners' vocabulary size or vocabulary depth could explain variance in reading comprehension. A third main aim for this study was to find out whether the frequency of occurrence of target non-words in a text (i.e., target words appeared either four or eight times in the passages), or the frequency band (high or low) of the original words which were replaced by the target non-words could explain learning and retention of these words. An interview was conducted to evaluate participants' familiarity with the target non-words. Finally, L1 English native speakers were compared with L2 learners of English in terms of word learning and retention of the target non-words, to find out if there were extraneous factors relating to the target words or the passages that could adversely affect the performance of the L2 learners.

Study 2 was conducted after obtaining the results of Study 1, in order to enhance L2 learning and retention of the meaning of new words encountered in a reading task. The study is based on Laufer and Hulstijn's (2001) Involvement Load Hypothesis (ILH). It first investigated the relationship between administering a task which requires learners to use a high level of *Evaluation* and L2 retention of the meaning of new words encountered in a reading task. The main objective in this study was the link between the ILH and L2 retention of the meaning of

new words, a relationship that may shed more light on the claimed predictive power of the Involvement Load Hypothesis. There is some evidence that strong *Evaluation* can lead to a higher rate of involvement in processing a new word and then lead to better learning and retention of this word (Keating, 2008; Kim, 2008; 2011). However, it is important to investigate further how L2 learners can be encouraged to engage more with the meaning of words in order to improve their learning and retention of the meanings of these new words.

Based on Jiang's (2000) model, the second aim of Study 2 was to investigate to what extent cross-language semantic priming enhanced word recognition in L2 learners of English, i.e., whether participants continue to map the L1 lemma content onto the target L2 form or whether they have started to build lemmas with L2 word meanings. It also aimed to examine whether the group which had carried out an elaboration task with these target non-words would succeeed in obtaining a priming effect from L2 semantically related prime word. Another question that needs to be answered was whether the task itself is the important factor, or the number of exposures the learner needs in order to learn the target words (Hulstijn & Laufer, 2001). Study 2 then aimed to reveal whether new words which replaced high frequency words or occurred more frequently in the text would be learned and retained better.

1.4 Statement of the Difficulties of Arabic Learners of English

Because the present research is in the context of learning vocabulary from reading, the discussion in this section will cover these same two areas of L2 vocabulary learning and reading. Vocabulary learning is consistently considered an important and essential element in L2 acquisition (Schmitt, 2000; Nation, 2001). This concept has been reiterated by Schmitt (2010), who argued that the importance of learning vocabulary is being noticed by learners themselves, who seek dictionaries more than grammar books. However, from my experience as an English language teacher at Qassim University, one of the L2 learners' challenges in learning English was related to their lack of English vocabulary. They also struggled to recall the meanings of even the most frequent English words, and this may be due to the way these words are taught. Nation (1990) emphasised the importance of teaching vocabulary in a more effective way to avoid difficulties that may occur in learning new words. According to my observation, students face difficulties not only in joining in an English conversation, but also in reading and understanding some of the crucial information provided by their teachers. This may occur if the English language classroom is the only source of input for L1 Arabic learners of English; access to other supplementary resources for developing their English language skills may be limited.

Based on my teaching experience, reading comprehension tasks are mainly designed to improve L2 learners' comprehension skills, rather than to develop their vocabulary knowledge. Students also experience limited reading practice in English, as the two most common teaching methods applied in reading classes are either the teacher reading a text aloud to the students or the teacher asking one of the students to read the whole text. Thus, the teacher's role in reading classes is to check the students' general comprehension of the text, ask some comprehension questions, and help the students translate difficult or unknown words in the passage into Arabic. It is also observed that most students hesitate and become stressed when reading or speaking in English, possibly due to the lack of practice they obtain in the target language outside the classroom; often they are not motivated to participate in the English Second Language (ESL) classroom.

Another challenge that may appear in reading classes is that little attention is given to students' preferences in choosing reading materials and topics, which limits students' enjoyment and increases their boredom in their reading classes. As a result, these traditional teaching methods in L2 reading classes seem to be ineffective. Furthermore, although selecting various topics to read may be more engaging and exciting for many of the students, it is common in most reading classes for students to read one long story, such as *Pride and Prejudice* by Jane Austen, for the entire academic year. It is, therefore, difficult for L2 learners to improve their reading skills or enrich their vocabulary knowledge from such reading activities. In addition, even if a few students seem to make a habit of reading in English, they often appear to suffer if they intend to practise reading on their own, due to the lack of appropriate English materials.

Research in the field of L2 vocabulary learning has emphasised the importance of finding the most effective teaching approaches in order to improve L2 learners' vocabulary. Hunt and Beglar (1998) suggested three approaches to enhance vocabulary learning: incidental vocabulary learning, explicit teaching, and independent strategy use, which can promote vocabulary learning. Incidental vocabulary learning seems to be an efficient approach in L2 acquisition (Hong, 2010). Many researchers would agree that incidental vocabulary learning is possible through spoken language (Macaro, Guo, Chen & Tian, 2009; Hennebry, Rogers, Macaro & Murphy, 2013) or via reading (Waring & Takaki, 2003; Horst, 2005; Pigada & Schmitt, 2006; Brown et al., 2008). Incidental vocabulary learning from reading is considered an effective approach to increase one's L2 vocabulary, because, as clarified by Huckin and Coady (1999), it involves reading and learning new words incidentally, and these two language learning skills can happen concurrently. It is also clear that L2 learners need support to enhance

incidental vocabulary learning from reading by the provision of, for example, dictionaries or glossaries (Knight, 1994; Hulstijn et al., 1996).

For these challenges faced by L1 Arabic learners of English in Saudi Arabia, I strongly suggest that such factors need to be taken into account in addition to the main teaching methods adopted by ESL teachers; teaching materials and activities that are appropriate for their proficiency level and suitable from the perspective of motivation should be used in the classroom. These factors may play an important role in students' academic success and in improving their English language skills.

In summary, as one of the researcher's motivations for choosing this topic for a PhD project, it is of interest to investigate the relationship between reading comprehension and vocabulary learning in order to, hopefully, make recommendations for the improvement of English language teaching and learning in Saudi Arabia that are based on research evidence. Findings from the current project can raise awareness of the significance of learning vocabulary from reading as a way to develop L2 lexical knowledge of L1 Arabic learners of English. Enhancement of L2 reading materials and activities is believed to provide better quality English language input (Nation, 1997), which may facilitate L2 learners' incidental vocabulary learning, and, as a result, help them to enlarge their vocabulary size and expand the depth of their knowledge of the words they have learned.

1.5 Structure of the Thesis

The current thesis consists of the following five chapters:

Chapter 1 is an introduction chapter. It provides background knowledge relevant to the present study and states the rationale as well as the aims of the current project. It also briefly discusses some of the challenges that L1 Arabic learners of English may face in learning the English language as a second language, particularly in L2 reading classes.

Chapter 2 reviews the studies relevant to the present topic. It first sheds lights on the constructs of reading comprehension, followed by the role of L1 in L2 reading comprehension. It then explores an important area of vocabulary knowledge, including defining a word, and discussing different aspects of vocabulary knowledge: organisation of words in the L2 mental lexicon, receptive and productive of word knowledge, size and depth of word knowledge. Next, it discusses the incidental and intentional learning approaches, followed by the vocabulary size requirement for language use in L1 and L2. The relationship between vocabulary knowledge and reading comprehension is discussed, and the forgetting curve is outlined. Hypotheses, such

as the Depth of Processing Hypothesis and the Involvement Load Hypothesis, are also covered in this chapter. Next, it describes the challenges of reading comprehension performance for L2 learners and then testing vocabulary knowledge is reviewed. The chapter concludes with a summary highlighting the most relevant key findings for the current study.

Chapter 3 presents the first part of the current project, Study 1. It contains four main sections: the aims of the study, the methods, the analysis, and a discussion of results. The first section begins with the provision of information about the rationale and aims for conducting the research of Study 1, followed by the research questions and the predictions addressed in the study. Next, it provides a description of the methods of the study, starting with information about the volunteers who participated in this experiment, followed by an overall design of the study. It also gives a detailed description of the research instruments used in this project: the Vocabulary Size Test (VST), the York Assessment of Reading for Comprehension (YARC), a Lexical Decision Task (LDT) that was developed for this study, and the Vocabulary Knowledge Scale (VKS). It then explains the procedures followed in the pilot study. The study procedures and the scoring procedures used to carry out this experimental research are also explained in this section.

The third section in this chapter presents the analysis of the results of Study 1. It begins with an explanation of the procedures followed in determining the results of this study. It further describes the statistical analysis of the data, which proceeds to answer the research questions raised by the study. The fourth section discusses the theoretical implications, starting with the key findings from Study 1, followed by further discussion of these findings in the light of the extant research presented in the literature review chapter. Lastly, this section offers some pedagogical implications of the findings that are relevant to the study's area, L2 vocabulary teaching and learning.

Chapter 4 presents the second part of the current project, Study 2. It consists of four main sections: the aims of the study, the methods, an analysis of results, and a discussion. The first section gives a description of the rationale and aims of the research of Study 2, highlighting the research questions and predictions addressed in this experiment.

The next section presents the methods used to conduct the study, beginning with information about the samples, and an explanation of the overall design of the study. It then describes an elaboration task (ET) which was developed for this study as a post-reading activity, followed by an explanation of the research instruments utilised in this experiment: the York

Assessment of Reading for Comprehension (YARC), Semantic Priming Task (SPT), Vocabulary Knowledge Scale (VKS), and Digit Span Test (DST). This section next gives a detailed description of the research design and procedures used in this project.

The third section in this chapter represents the results of the data analysis. It begins with a description of the procedures, followed by the results of Study 2. It presents the statistical analysis of the data to answer the research questions raised by the study. The purpose of the final section is to discuss the results of Study 2, presenting the key findings of the study, and then examining the theoretical implications of the study's results in light of the existing research included in the literature review chapter. Lastly, this section discusses the pedagogical implications of the findings that should be considered when designing and using L2 vocabulary activities.

Chapter 5 presents the conclusion of the studies. It shows how the findings from both studies can contribute to the field's knowledge. This chapter also discusses the limitations of the studies and provides suggestions for future research.

Chapter 2: Literature Review

2.1 Overview

Reading has always been regarded as an important skill for academic success, both in first language (L1) and second language (L2) learning (Johns, 1981). Developing a large enough vocabulary is an essential element in L2 acquisition to be able to read and write and participate in various topics of conversation. One of the primary factors consistently shown to affect reading comprehension is knowledge of the words in the text (Laufer, 1992). Researchers have attempted to define the percentage of vocabulary needed by a second language learner in order to be able to read adequately (Laufer & Sim, 1985; Laufer, 1989; Hsueh-chao & Nation, 2000; Schmitt et al., 2011). According to Hsueh-chao and Nation (2000), learners need to know at least 98% of the words for comprehension of written texts, which requires knowledge of 8,000 to 9000 word families (Nation, 2006). For taking part in daily conversations, it is assumed learners need about 2000 - 3000 word families (Adolphs & Schmitt, 2003). One of the issues that remains controversial is how learners acquire the vocabulary needed in everyday life (Hill & Laufer, 2003).

Researchers have indicated the relevance of reading in acquiring vocabulary incidentally, as a by-product of reading (Horst et al., 1998; Waring & Takaki, 2003; Horst, 2005; Pigada & Schmitt, 2006; Brown et al., 2008), but vocabulary gains are often rather limited by approximately one new word in every five words (Horst et al., 1998). L2 learners face difficulties in learning thousands of words, particularly in contexts where time and resources that can be used for language support are limited, as in the case of many Arabic learners of English (Akasha, 2013). Teachers might know about their students' difficulties but might struggle to find the most effective way to help students enrich their L2 vocabulary. Therefore, there is clearly a need for research that can help to identify effective ways that provide optimal chances for L2 vocabulary learning.

As discussed in the previous paragraph, there is thus a bidirectional relationship between vocabulary and reading: vocabulary knowledge affects reading comprehension, and, conversely, reading comprehension affects vocabulary learning. Reading comprehension research also confirms that extensive repetition of word forms contributes greatly to acquisition of new vocabulary (Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010). These findings are therefore extremely important for designing new methods to develop learners' reading skills and to enrich their vocabulary.

A number of researchers have offered different approaches to enhance vocabulary learning. As an example, Hunt and Beglar (1998) suggested that incidental vocabulary learning (that is, learning words as a by-product of reading), explicit teaching of words, and independent strategy use can promote vocabulary learning. While many researchers would agree that incidental vocabulary learning is possible, it is also clear that L2 learners need support to enhance incidental vocabulary learning, for example, in the use of dictionaries or providing glosses (Knight, 1994; Hulstijn et al., 1996). Laufer and Hulstijn (2001) operationalised the Depth of Processing Hypothesis by proposing the ILH. The cited authors argue that L2 incidental learning and retention of new words depends on the amount of involvement load that a task induces and the degree of depth of processing needed to complete that task. Laufer and Hulstijn (2001) and Kim (2011) also call for more empirical research in order to find out the effectiveness of using vocabulary tasks with different involvement loads on L2 vocabulary learning.

This chapter provides an integrated review of the key studies relevant to the present study. This review of literature covers two main themes: the construct of reading comprehension and L2 vocabulary acquisition. The first part of the chapter sheds light on defining the construct of reading comprehension and the role of L1 in L2 reading. Then, different aspects of vocabulary knowledge are discussed: the organisation of words in the L2 mental lexicon, receptive and productive of word knowledge, size and depth of word knowledge. This chapter further investigates incidental and intentional learning approaches and the vocabulary knowledge that is required for reading. In order to investigate the relationship between vocabulary knowledge and reading comprehension, a number of existing publications in the SLA area are reviewed. It also offers a review of the main related theories, the Depth of Processing Hypothesis and the Involvement Load Hypothesis. It furthermore describes the challenges of reading comprehension performance for L2 learners who use a different script in their L1, as is the case for Arab learners of English language. Lastly, this chapter provides a conclusion and summary presenting the key findings from the literature, as relevant for the current project.

2.2 Defining the Construct of Reading Comprehension

Before we can discuss to what extent L2 learners learn words from reading, it is important to define the construct of reading comprehension. Reading comprehension is defined as "the process of construction of meaning through the dynamic interaction between the reader's existing knowledge, the information suggested by written language, and the context of the reading situation" (Anthony, Pearson & Raphael, 1993, p. 284). Similarly, Schelling, Aarnoutse

and Leeuwe (2006) define reading as an interactive activity between the reader's knowledge and the characteristics of the written input such as genre and structure. Reading involves different skills, including skimming for the main idea, scanning for specific details, interpreting difficult meaning and guessing the meaning of unknown words (Hughes, 2003). Brown (2004) categorised reading skills into micro-skills and macro-skills. The micro-skills involve how to recognise the word and understand the meaning in different structures, while the macro-skills refer to understanding the main idea, making links between actions, and applying reading strategies such as skimming and scanning or identifying discourse makers.

As the focus of the present project is on reading comprehension, it is important to mention a theoretical framework that is widely used in the field, namely, the Simple View of Reading (SVR). Originally proposed by Gough and Tunmer (1986), and supported by Hoover and Gough (1990), the SVR proposes that reading comprehension can be predicted by two main components: individual's decoding and linguistic comprehension skills. According to the model, reading comprehension is defined by the formula:

Reading comprehension = Decoding x Linguistic Comprehension

For the purpose of the SVR, decoding and linguistic comprehension are defined as follows:

Decoding means matching letters of words with their correspondent sounds to read them accurately or the ability to recognise the word. It is also referred to as efficient word recognition: the ability to read familiar and unfamiliar words with accuracy and speed from a given text or list.

Linguistic comprehension is defined as the ability to process and understand the spoken language. It involves a number of language skills, such as the ability to derive lexical semantic information, the representation of sentences, and the ability to categorise and process sounds. Reading comprehension "involves the same ability, but one that relies on graphic-based information arriving through the eye" (Hoover & Gough, 1990, p. 131).

Both decoding and linguistic comprehension are essential for reading comprehension. Weakness in either decoding or linguistic comprehension skills, or both skills, would lead to weakness in reading comprehension skills. Studies based on this model propose that decoding and language comprehension account for between 40% and 70% of the variance in reading comprehension (Cutting & Scarborough, 2006; Joshi & Aaron, 2000).

Many authors have focused on the influence of decoding and linguistic comprehension on L1 reading comprehension in the framework of the SVR (Chen & Vellutino, 1997; Joshi & Aaron, 2000; Adolf, Catts & Little, 2006). Other scholars suggest that the SVR framework explains not only L1 reading comprehension, but also L2 reading comprehension (Droop & Verhoeven, 2003). The SVR model is also examined at different ages, e.g., in elementary students (Verhoeven & van Leeuwe, 2008), and adults (Sabatini, Sawaki, Shore, & Scarborough, 2010; Braze, Katz, Magnuson, Mencl, Tabor, Van Dyke, Gong, Johns, & Shankweiler, 2016).

While the SVR model provides a clear insight into the main components required for skilled reading comprehension, Tunmer and Chapman (2012) reviewed the SVR to find out whether vocabulary makes an independent contribution to variance in reading comprehension. They suggest that this is indeed the case and that vocabulary knowledge explains variance in reading comprehension over and above the contribution of decoding and linguistic comprehension. However, the authors support the original two-component structure of the SVR: vocabulary is not a separate component of reading alongside decoding and linguistic comprehension. Instead, they suggest that vocabulary knowledge is a component of linguistic comprehension and that linguistic comprehension contributes to reading comprehension both directly and indirectly via decoding. The indirect contribution of linguistic comprehension is indicated in Figure 2.1 through the addition of a path from linguistic comprehension to decoding (see Figure 2.1).

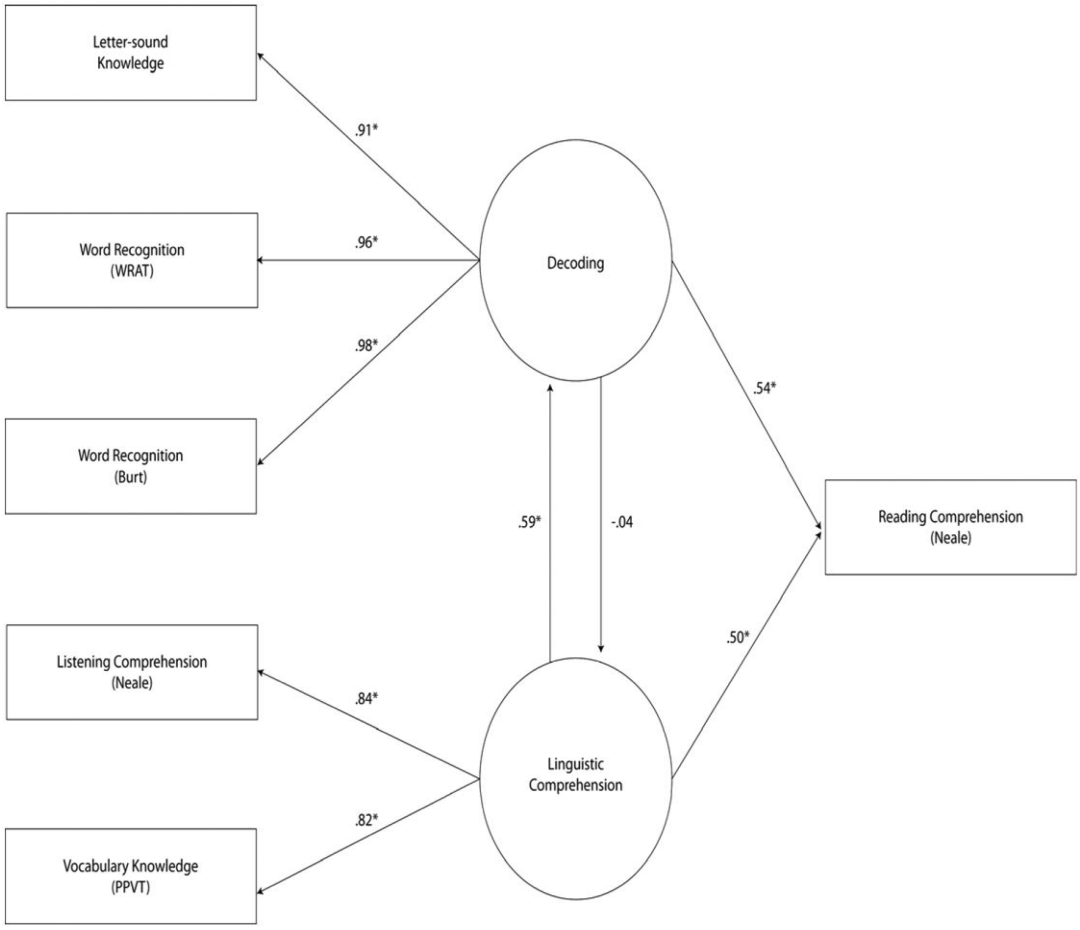


Figure 2.1 Modified Sample View of Reading Model (Tunmer & Chapman, 2012)

Although reading in adult samples has not been studied in as much detail as those of younger readers in primary schools (Curtis, 2002), the SVR model has successfully been applied to adults (Sabatini et al., 2010; Braze, Tabor, Shankweiler & Mencl, 2007; Braze et al., 2016). In a replication of Tunmer and Chapman's (2012) study, Braze et al. (2016) examined reading in an adult sample and found a significant contribution of vocabulary knowledge to reading comprehension. They support Tunmer and Chapman's (2012) view by finding that vocabulary knowledge is subsumed under the linguistic comprehension component. Therefore, the studies reviewed here illustrate the importance role of vocabulary knowledge for reading comprehension, although the number of studies working with this framework on adult L2 learners is very limited.

2.3 The Role of L1 in L2 Reading

In terms of L2 reading, learners tend to use their L1 as a medium tool in order to recognise and understand the concept of the written input (Grabe, 1991). Most of earlier research on linguistic transfer has mainly concentrated on the role of L1 in decoding during L2 reading (Durgunog˘lu, Nagy, & Hancin-Bhatt, 1993). Researchers (Durgunog˘lu, et al.,1993; Grabe,

2009; Koda, 2004) have pointed out that L2 learners have the ability to develop word decoding skills rapidly in the target language. They assume that this may be attributed to the L1, in the case that learners can read in their L1. It means that being able to read in L1 might have a positive effect on achieving success in L2 phonology and orthography, which may help them develop L2 decoding abilities quickly. Furthermore, researchers have found that several key elements that are related to reading comprehension transfer across languages. For instance, higher order word knowledge, particularly the capacity to define vocabulary, was found to be associated across languages (Ordóñez, Carlo, Snow, & McLaughlin, 2002).

In terms of the role of L1 reading, earlier studies have investigated the impact of L2 proficiency and L1 reading ability on L2 reading comprehension (Clarke, 1979; Cziko, 1980; Devine, 1987), and have argued that L2 reading ability is largely dependent on learners' proficiency in that language has less correlation with L1 reading ability. Clarke (1979) conducted a study on L2 (English) learners' ability to transfer L1 reading (Spanish) to L2 reading. The findings revealed that L1 reading strategies were not transferred to L2 because of the limitations of proficiency in L2. Similarly, Ulijn and Kempen (1976), who investigated Dutch students reading French, concluded that L2 comprehension is not dependent on the contrasts between L1 and L2, rather, it depends on the conceptual knowledge of the learner in L2.

More recent studies have also provided evidence that L2 proficiency is one of the most effective factors with regard to L2 reading comprehension. In investigating the interrelationships of L1 literacy, L2 proficiency, and L2 reading comprehension, Jiang (2011) conducted a study with 246 Chinese college learners of English. The results indicate that L2 language proficiency accounted for 27% - 39% of variance in L2 reading comprehension, whereas L1 literacy accounted for less than 6% of the variance. This finding confirms the assumption that L2 language proficiency results in a higher contribution to L2 reading comprehension performance. According to Jeon and Yamashita (2014), there are strong correlations between components of L2 knowledge and reading. They found that L2 grammar knowledge (r= 0.85), L2 vocabulary knowledge (r= 0.79), and L2 decoding (r= 0.56) obtained the three strongest correlations with L2 reading comprehension, whereas L1 reading comprehension (r=0.50) obtained a lower (medium) correlation with L2 reading comprehension. Similarly, other studies in various languages with L2 learners have found evidence that L2 language proficiency was a stronger predictor of L2 reading comprehension (Tsai, Ernst & Talley, 2010; Guo & Roehrig, 2011).

Thus, there is some evidence L1 decoding has a positive significant influence on achieving success in L2 orthography features, which may lead to developing L2 decoding abilities quickly. For L2 reading comprehension, however, it appears to be the case that that L2 proficiency has more impact on L2 reading comprehension than L1 reading ability. As such, L1 reading ability was not taken into consideration in the current study as a factor that may affect L2 reading comprehension assessment.

2.4 Vocabulary Knowledge

2.4.1 Knowing a word

A simple question, such as 'What does it mean to know a word?' can have several meanings, depending on how the listener interprets it. L1 and L2 lexical researchers have made various proposals for understanding what it means to know a word (Richards, 1976; Nation 1990; 2001; Qian, 1998). Richards (1976) was one of the early studies on vocabulary acquisition which formulate a framework for describing what is included in knowing a word. According to him, it involves knowing:

- "the degree of probability of encountering the word in speech or print" (p. 79), e.g., *book* is more frequent than *manual.*

- "the limitations imposed on the use of the word according to the function and situation" (p. 79), like the British use the word *tap* and in American English the same object is called a *faucet.*

- "the syntactic behaviour associated with the word" (p. 80), e.g., using the term *structural words* for a number of frequent words in the vocabulary.

- "the underlying form of a word and the derivations that can be made from it" (p. 80), such as words with regular derivations, e.g., *walked, walking, walks* are easily recognised as derived from *walk.*

- "the associations between the word and other words in the language" (p. 81), e.g., *old* is linked to *new,* and *good* to *bad.*

- "the semantic value of a word" (p. 82), i.e., the semantic features associated with the word, such as a *man* consists of *human + male.*

- "the different meanings associated with the word" (p. 82), such as *shop, store,* and *boutique.*

Richard's paper is still highly influential in the field of vocabulary acquisition (e.g., Nation, 1990; Schmitt & Meara, 1997). However, these assumptions have been criticised by Qian (2002), as not including the spelling and pronunciation aspects. Similarity, Meara (1996) clarifies that Richard's (1976) framework does not seem to provide an appropriate theoretical model for word knowledge, but it can be suitable for classroom instruction. Nation (1990) maintains that most learners have partial knowledge of different aspects of word knowledge and that learners differ from each other in how well they know these.

Nation (2001) provides a more comprehensive model for defining word knowledge, covering all aspects of what is included in knowing a word. He states that vocabulary knowledge involves two dimensions: receptive knowledge, which refers to how learners understand the words when listening or reading, and productive knowledge, which refers to how learners use words when speaking or writing. Nation (2001) divides word knowledge into three main categories: knowledge of *form*, knowledge of *meaning*, and knowledge of *use.* Each of these three can be receptive and/or productive (see Table 2.1).

In Table 2.1, Nation categorises receptive knowledge (shown as R in Table 2.1) and productive (shown as P in Table 2.1). Nation maps receptive and productive distinctions to the subdivisions, which means the form/spoken/receptive division aims at answering the question "What does the word sound like?" and the form/spoken/productive division clarifies "How is a word pronounced?" According to Nation's categorisation, knowing the *form* of a word involves knowing its spoken and written form, and its parts, whereas knowing the *meaning* requires knowledge of the link between a word's form and meaning, and knowing the link between the concept of meaning and referents, as well as knowledge of word associations. Knowledge of word *use* requires knowing its grammatical functions, collocations, and the constraints on word use. With respect to word knowledge, vocabulary size involves the *form* as well as the sub-category *form and meaning* (Milton, 2009). While vocabulary depth involves all aspects mentioned by Nation (see Table 2.1), depth of word knowledge relates to how well the different aspects of word knowledge are known. The distinction between size and depth is not always clear. This issue will be further explained in Sections 2.4.2.3 and 2.5.2.

Table 2.1 What is involved in knowing a word.

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognisable in this word?
		P	What words parts are needed to express meaning?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concepts and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What others words does this word make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of word occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use	R	Where, when and how often would we meet this word?
		P	Where, when and how often can we use this word?

Source: Nation (2001)

All of these aspects of words are closely related to each other and are connected. This understanding of the lexical aspects approach poses the question as to how much and which aspect of word knowledge is most important for learners. This may depend on the word class itself and on the task requirements. As an example, knowledge of verbal inflections is an essential element in recognising and being able to produce a verb. For adjectives, however, the semantic knowledge and its relation of gradation are crucial elements in acquiring the word meaning (Henriksen, 1999). At the same time, form-meaning links of the lexical item are considered to be the most important word aspect that must be required, and probably enough to allow recognition.

According to Alderson (2000), the definition of the word should include knowing its meaning and knowing how to use it accurately in context. Nation (2001) claims that knowing a word involves focusing on the word-system and patterns which provide an explanatory view of that word. This will help the learning burden to be very light, making it easier to learn the word. The learning burden means the amount of effort needed to learn a word, and each aspect of word knowledge can contribute to a word's learning burden (Vela, 2014). The general concept of learning burden, according to Nation (1990), is that the more patterns and knowledge about a word are already known by learners, the lighter its learning burden. Vela (2014) claims

that a word is generally expected to be easy to learn if it a) uses sounds that are in the learners' L1, b) has regular spelling patterns, c) shares with the L1 roughly the same meaning, d) has roughly similar grammatical patterns to the L1 and fits into similar collocations and constraints. In other words, for L2 learners whose first language is quite similar to the second language, the learning burden of most words seems to be light, which means they would not be difficult to learn. On the other hand, the learning burden is assumed to be higher for L2 learners whose first language is not closely related to their second language (Vela, 2014).

It is assumed that some of the word knowledge aspects are established sooner than others, due to the different stages of word acquisition (Schmitt, 1998). Schmitt (2010) also suggested that at the beginning of the learning process, it seems better to focus on mastering the meaning-form link. With further word mastery, it may be more effective to enhance other aspects of word knowledge, such as collocations (Schmitt, 2010). Nevertheless, vocabulary researchers may not be able to measure all of these aspects in any test. This is because such a test would be too long and complex. Thus, researchers must be very careful in choosing which dimension they are going to test in their vocabulary studies and consider the boundaries of their choices (Schmitt, 2010).

2.4.2 The basic dimensions of word knowledge

L2 vocabulary acquisition researchers provided a different, but complementary, framework for defining what it means to know a word (Read, 2000; Nation, 2001; Qian, 2002). Qian (2002) argued that four basic dimensions should be considered in analysing what vocabulary knowledge means: lexical organisation, receptive-productive knowledge, size, and depth of vocabulary. L2 lexical organisation research show how L2 words are processed and represented in the L2 mental lexicon of a learner. The receptive-productive knowledge dimension refers to the processes involved in recognising or producing words, while vocabulary size and depth refer to how many words are known and to how well these words are known. Considering the importance of L2 lexical organisation, it is necessary to see how the lexical item is processed and represented in the L2 mental lexicon of a learner. Thus, before discussing the receptive-productive knowledge, size and depth of vocabulary dimensions, an insight into the organisation of words in the L2 mental lexicon is provided.

2.4.2.1 Organisation of words in the L2 mental lexicon

To understand L2 vocabulary acquisition, it is necessary to answer the question of how a lexical item is organised in the L2 mental lexicon. Jiang (2000) proposes a psycholinguistic

model of lexical representation and development. As the current study is based on Jiang's (2000) model, it will be further explained in this section.

Jiang (2000) argues that adult learners who have already established their L1 lexical system and are learning the L2 in an instructional context may have limited access and exposure to the L2. Jiang (2000) points out that the process of L2 vocabulary acquisition comprises three stages. The first stage is the *formal stage*. At this stage, a lexical entry associated with formal specifications, i.e., phonological or orthographic information, is established for an L2 word, but this lexical entry is not yet complete. It consists of a "pointer" that links an L2 word to its L1 lemma. When this L1 word is accessed, the morphological information as well as the semantic and syntactic information that is stored in L1 lemma become available (see Figure 2.2).

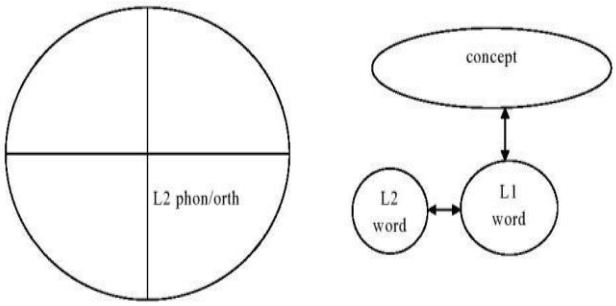


Figure 2.2 Lexical representation (left) and processing (right) in the formal stage (Jiang, 2000, p. 51)

In the second stage, the connection between L2 words and L1 lemma (semantic and syntactic information) is gaining strength due to the continuous simultaneous activation. Therefore, Jiang proposes that at this stage the lexeme would contain L2 formal features, while the lemma would continue to contain the L1 semantic and syntactic information. This second stage is called the *L1 mediation stage* because the L1 lemma features are transferred into the L2 lexeme and therefore the L1 mediates L2 word use. The L2 semantic information is missing at this stage. Furthermore, Jiang (2000) suggests that the relationship between the L2 word form and the L2 concepts is not so strong at this stage, since the lemma information is copied from the L1 during the process of acquiring the L2 word, which makes the integration of this information quite weak (see Figure 2.3).

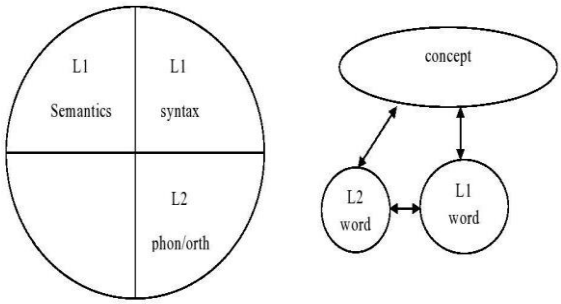


Figure 2.3 Lexical representation (left) and processing (right) in the L1 lemma mediation stage (Jiang, 2000, p. 53)

The last stage is called the *L2 integration stage*. Receiving adequate L2 input and practice results in the gradual extraction of the L2 semantic syntactic information, and this is then integrated into the lexical entry to replace the L1 lemma information (see Figure 2.4).

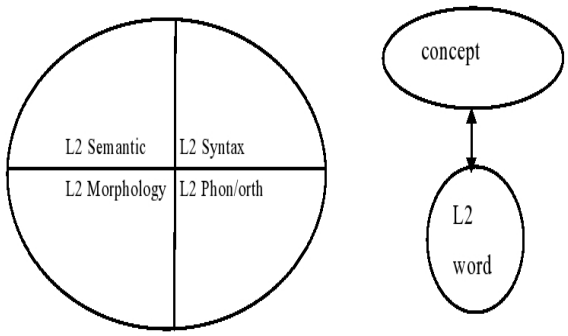


Figure 2.4 Lexical representation (left) and processing (right) in the L2 integration stage (Jiang, 2000, p. 53)

Jiang (2000) clarifies that words may be at different stages in the development process. This may result in differences in using L1 lemma information in production and comprehension. Jiang also observes that even under the most desirable learning conditions with an adequate L2 input, sometimes learners are unable to extract L2 lemma information from this input. Jiang claims that most of the L2 words fossilise at the L1 lemma mediation stage, because the presence of the previously established L1 lemma may prevent L2 learners from extracting L2 lemma information from the input. Jiang (2000) suggest that it is the L1 lemma in the L2 lexical entry which blocks access to L2 semantic and syntactic information. As a result, the transfer from the L1 mediation stage to the L2 integration stage may therefore never succeed (Jiang, 2000).

Cui (2009) also suggested that in order to achieve success in L2 vocabulary acquisition, L2 learners have to establish a semantic system that is specifically for the L2 vocabulary and

completely free from the effect of the L1 semantic system. Contradictory to this finding, Jiang (2002) found that L1 semantic content is still present in the L2 lexical entries, even with advanced L2 learners. In Jiang's (2002) study, English native speakers and L1 Chinese speakers of English took part in two experiments. In particular, two semantic judgment tasks were conducted in which participants were asked to make a judgment on the degree of semantic relatedness of two English words (experiment 1) or to decide whether English word pairs were semantically related or not (experiment 2). The results of both experiments offer evidence for the presence of L1 semantic features in L2 lexical entries. It was found that non-native speakers obtained higher scores and shorter times in responding to English word pairs sharing the same L1 semantic information than to the different translation pairs, while no differences were obtained between these two types of L2 words pairs for English native speakers. This result was also confirmed by Jiang's study (2004) in a replication of the second experiment (Jiang, 2002) with a different ESL population, L1 Korean speakers of English as an L2. The results also provide some evidence for L1 mediation in L2 lexical processing. Furthermore, since all participants in both studies had high L2 proficiency levels, these findings confirm that L1 semantic features are still present, even with advanced L2 speakers.

Jiang (2004, p. 428) clarifies that it is still not clear whether the same L1 translation effect can be found with speakers whose L1 is "typologically related to English, such as Spanish or German." Jiang (2000) also confirms that some adults in instructional contexts can achieve the last stage of L2 development, whereas others, even in naturalistic contexts their L2 lexical entry, may fossilise at the second stage. This means that factors other than input and the presence of L1 lemma information in the L2 lexical entry may affect L2 lexical processing and development. Moreover, it is important to know that individual words can go through these three stages at different rates.

Lexical transfer means, according to Jiang (2000), "the use of L2 words on the basis of their L1 translations" (p. 66). Jiang (2004) notes that L2 learners can only approximate the L2 lexical competence of native speakers if L2 semantic content is integrated into the L2 lexical entries of words. Overall, Jiang's (2000) model provides an important insight into L2 lexical processing and development which can help to better understand L2 vocabulary acquisition. Furthermore, as Jiang clarifies, the fact that L1 semantic content continues to play a role in L2 lexical processing, even with advanced learners, has important practical implications for language teaching, which will be discussed in Chapter 4.

2.4.2.2 Receptive and productive vocabulary knowledge

One of the important distinctions in the field of vocabulary acquisition is the division of word knowledge into *receptive/passive* knowledge and *productive/active* knowledge (Milton, 2009). Despite the number of researchers and language teachers who provide definitions and distinction, these two notions are not easy to identify (Melka, 1997; Laufer & Paribakht, 1998). The general definition is that receptive knowledge of a word refers to the ability to realise a word's form while listening or reading, and to recall its meaning, whereas productive knowledge means the ability to produce a spoken or written form of a word and the ability to express the meaning of a word through speech production or writing (Nation, 2001; Milton, 2009). In other words, when L2 learners become aware of a word presented in a text or heard in a conversation, but they cannot really produce it, then this is simply called a receptive word knowledge. The productive knowledge of a word, however, requires L2 learners to use the word productively through speaking or writing. With respect to the multi-aspect of word knowledge, receptive vocabulary knowledge means being able to recall and recognise aspects of word knowledge in reading or listening, whereas productive vocabulary knowledge refers to the ability to use aspects of word knowledge in writing or speaking (Zhong, 2016).

Schmitt (2010, p. 82) argues that, due to the complexity of the nature of word knowledge, receptive and productive knowledge cannot be acquired in "a uniform manner." Nevertheless,the distinction between receptive and productive knowledge is not very clear. Melka (1997, p. 87) describes this distinction as "a continuum," degrees of knowledge, rather than two separate systems, and different aspects of word knowledge are possibly known to different degrees receptively and/or productively.

Therefore, the basic points to be taken from the discussion of these two dimensions are that it is often assumed that receptive knowledge precedes productive knowledge. L2 learners first acquire a word receptively, and then through further exposure to this word, their knowledge may gradually improve to a productive level of use (Melka-Teichroew, 1982; Melka, 1997). The receptive level of word knowledge can develop to a productive level (Milton, 2009), however, it is not clear yet why productive learning and use are more difficult than receptive learning and use. One of the possible explanations is that productive knowledge requires more precise knowledge of the word than receptive knowledge (Nation, 2001). There are various competing words, thoughts need to choose from in order to be produced, and in the case for languages as claimed, the ones that are more equivalent with the L1 lexical system seem to be stronger (Ellis & Beaton; cited in Nation, 2001).

Another related assumption is that learners' receptive vocabulary is larger than their productive vocabulary (Melka, 1997). It is estimated that 50-80% of the words which a learner has learned at a receptive level of knowledge are also used productively (Milton, 2009). On the other hand, Webb (2005) claims that receptive and productive knowledge are two distinctive dimensions. That is, new words acquired receptively will lead a learner to gain more receptive knowledge, whereas productive learning of words will also enhance larger productive knowledge.

Last but not least, since the distinction between the receptive and productive knowledge of a word can provide an insight into a scale of knowledge and identify the level of word knowledge, vocabulary acquisition research should take these perspectives into consideration and the results should be interpreted based on these two dimensions.

2.4.2.3 Knowledge: Size and depth

Many researchers assume that vocabulary knowledge includes the following two primary dimensions: size and depth (Wesche & Paribakht, 1996; Qian, 1998). The term "vocabulary size" is generally used in the literature to replace the "vocabulary breadth" dimension (Meara, 1996). Both terms, "size" or "breadth," are used interchangeably in the literature to denote the same concept. In the current research, however, "vocabulary size" will be used to represent the vocabulary breadth of learners. Size of knowledge refers to the number of words a learner knows in a single meaning within a single context.

Depth of knowledge means an adequately deep understanding of a word (Schmitt, 2010). It links to the quality of word knowledge, or how well a learner understands a word (Milton, 2009). As Nassaji (2004, p. 112) points out, researchers have shown "the complexity and multi-dimensionality of word knowledge … have suggested that knowing a word well should mean more than knowing its individual meanings in particular contexts." Different approaches have been carried out in order to define depth of word knowledge. Some researchers consider depth of word knowledge as including the knowledge of a word's relationship to other words in the lexicon (Haastrup & Henriksen, 2000; Meara, 2006). Other researchers have argued that morphological information is an important aspect of word knowledge, as it can help to enrich the learner's word knowledge as well as the learner's understanding of the word's relationships to other words (Qian, 1999; Perfetti, 2007; Kieffer & Lesaux, 2008). Qian (1999; 2002) assumed that depth of vocabulary involves all features related to semantic, pragmatic, morphological, syntactic, collocational, and phonological knowledge of a word. Daller, Milton, and Treffers-Daller (2007) summarised the various aspects of word knowledge by using three

theoretical dimensions: size, depth, and fluency. Learners who are highly fluent can easily access the words they know without any hesitation, while learners who are less fluent have difficulty in recognising and knowing the words they are trying to use. Further details about these two dimensions will be explained in Section 2.5.2. In the following, the discussion will focus on the vocabulary size required for language use in both L1 and L2.

2.4.3 Vocabulary size in L1 and L2

2.4.3.1 Native speakers' vocabulary size requirement for language use

A critical question in vocabulary studies is how much vocabulary is necessary to allow for communication. Many studies with varying designs have been utilised to explore the vocabulary size of native speakers; these have led to widely varying estimates (Nation, 1993). In one of the earliest studies, Kirkpatrick (1891) estimated that US high school students had a vocabulary of approximately 10,000 words, while college students between 20,000 to 100,000 words. It has also been estimated that well-educated adult native speakers of the English language know about 17,200 word families in Webster's Third International Dictionary (1963) (Goulden, Nation & Read, 1990). Each word family comprises individual word forms, involving the root form, e.g., (*stimulate*), its inflected forms (*stimulated, stimulates and stimulating*), and regular derivations (*stimulative, stimulation).*

Zechmeister et al., (1995) found that university students know around 16,000 word families in the 1980 Oxford American Dictionary, while high school students know an average of 11,836 and retired adults know around 21,252 headwords (word families) consisting of all inflected and derived forms. Nation and Waring (1997) posited that children who are native speakers of English normally add about 1,000 word families per year to their vocabulary. This means that a five-year-old child begins school with a vocabulary of around 4,000 to 5,000 word families; this same child graduates from university with a vocabulary of about 20,000 word families. According to Schmitt (2010), a range of 16,000 - 20,000 word families is typical for educated native speakers.

In investigating how many words undergraduate English native speakers know, Milton (2009) proposed the smallest estimate of about 9,000 word families. Milton and Treffers-Daller (2013) also found that undergraduate monolingual English speakers' vocabulary size seems much smaller, about 10,000 English word families, than has been assumed previously, with their vocabulary size increasing gradually during their studying at university by 400 to 500 words per year. Native speakers, therefore, vary in their vocabulary size, depending on how much and how they use their language.

2.4.3.2 Second language (L2) learners' vocabulary size requirement for language use

Fortunately, non-native speakers of a particular language do not need access to a native-like vocabulary to use the language effectively (Schmitt, 2010). For example, Staehr (2009) concluded that advanced Danish listeners who learned the 5,000 most frequent word families in English were able to demonstrate adequate listening skills for the Cambridge-ESOL Proficiency Certificate in English (CPE) listening exam. Schmitt (2010) noted, however, that such data suggests that if 95% coverage (known words) is adequate, L2 learners need to know between 2,000 and 3,000 word families to be conversant in English; alternatively, if 98% coverage is required, knowledge of between 6,000 and 7,000 word families is needed. These estimates do not clearly indicate the coverage requirement for listening, as it is not yet clear whether the 98% coverage figure that is derived from written vocabulary research is the most appropriate figure for spoken discourse.

For estimates of written vocabulary, second language learners need to learn a large number of individual word forms to be able to read different types of texts in English. Nation (2006), based on the British National Corpus (BNC) data, concluded that 8,000 - 9,000 word families are necessary to read authentic texts such as novels or newspapers without unknown vocabulary being a problem. According to his calculation, a vocabulary of 8,000 word families requires knowing 34,660 individual word forms, including the family members of low-frequency items. In order to estimate the word families required for daily use, Nation (2006) investigated word lists of approximately 20,000 words, based on the British National Corpus (BNC), which involved radio talk, conversation (between family members and friends) and interviews. He found that about 6000 - 7000 word families are essential in order to be able to deal with English conversation, more specifically, to achieve the 98% goal.

As Nation (2006) points out, the most frequent 1,000 word families in the BNC word lists consist of approximately six types per family, declining to roughly three types per family at the 9,000 frequency level. Occasionally these word family types are clearly related, such as *nation* and *national*, which could be easier to guess. Others are less clearly related such as *nation* and *nationalistically*, which makes it more difficult for learners to recognise unknown words. Laufer (2000) reviewed a number of vocabulary studies from different countries and found that the vocabulary size of educated second language learners of English ranged from 1,000 - 4,000 word families. This size seems to be very small and it may not be enough for educated L2 learners of English. She also showed that a reading target of 3,000 - 5,000 word families can

be attainable for second language learners, however, the 8,000 - 9,000 targets can be difficult to achieve.

In addition to the importance of a large vocabulary for language use, expanding the word knowledge is also important. This is because knowing a great deal about each individual lexical item is important in order to use it accurately. Some theoretical concepts related to vocabulary knowledge, particularly L2 incidental vocabulary acquisition will be discussed in the following section.

2.4.4 Defining explicit/implicit and incidental/intentional vocabulary learning

The question of how L2 vocabulary knowledge is acquired is often discussed in the context of the distinction between explicit and implicit learning. Implicit learning, as defined by Ellis (1994a), is an automatic, natural acquisition of words which takes place without conscious operations, while explicit learning, on the other hand, is a conscious process, including searching for information, and formulating and testing hypotheses. In terms of vocabulary learning, Ellis (1994b) pointed out that the phonological features, such as grapheme-phoneme, and spoken word production, are acquired implicitly as a result of learning perceptual aspects of lexical items. Word forms are also another aspect of word knowledge which can be improved implicitly. The meaning of words, however, is acquired explicitly because it requires more conscious operations at the semantic level and involves making form-meaning connections.

It is also important to clarify that implicit and explicit learning should not be confused with the notions of incidental and intentional learning. Implicit learning, without learners being conscious that they are storing information in memory, can happen incidentally only, whereas explicit learning can be both incidental and intentional (Hulstijn & Laufer, 2001). Therefore, as making form-meaning connections does not require explicit learning, L2 vocabulary can be acquired incidentally and intentionally as well. The term “incidental vocabulary acquisition” is considered to mean learning without an intention to learn (Laufer & Hulstijn, 2001). It is defined as the process in which words are learned while focusing on the content of the message (Nation, 2001), for instance, learning the meaning of new words because of a reading task. In other words, incidental vocabulary acquisition means that any words are acquired as a by-product of an activity that may not target vocabulary learning directly (Gass, 1999). The effectiveness of incidental vocabulary acquisition is supported by Paribakht and Wesche (1997), who were among the first to clarify that vocabulary learning from reading can be enhanced through a reading task combined with exercises. They found that new lexical items practised through tasks were understood and retained better than words for which meanings

were inferred from a particular context. In a similar vein, Laufer and Hill (2000) pointed out that reading, listening, and speaking activities can enhance incidental vocabulary learning. Hennebry et al. (2013) pointed out that the degree of incidental learning of new words from reading depends on the amount of time specified to process that word while reading in order to understand the content of the text.

As Laufer and Hulstijn (2001) pointed out, in incidental learning, students are asked to perform a task involving processing of information of some kind without being told they will be tested on this information afterwards, whereas in intentional learning, students are told they will be tested afterwards. Barcroft (2009) proposed that incidental and intentional learning can be considered as the endpoints of a continuum, as learning processes can pass through a sequence scoping from more intentional to more incidental. Nation (2013a), however, clarified that these two approaches complement each other, which means combining them in a learning plan can be an effective means to facilitate vocabulary learning.

Elgort (2011) has found that deliberate vocabulary learning can enrich both explicit and implicit knowledge. In her study, she used three different kinds of priming tasks, namely, for priming, meaning priming, and repetition priming. The words used as primes in the three experiments were 48 pseudowords which were specifically created by changing one letter from the base word (original word) so that they became an unfamiliar word, for example, *maxidise* was created from the base word "maximise." These pseudowords had been studied using word cards one week before the experiment and were briefly shown on a screen prior to the appearance of the target words. A lexical decision task was used to test L2 learners' ability to distinguish between real words and pseudowords. The results indicated that the pseudowords were successfully processed as primes, which provides some evidence for the assumption that they had been transferred into the real lexical system. Elgort (2011) concludes that deliberate learning can therefore enrich L2 vocabulary acquisition.

Despite the different directions to define incidental vocabulary learning, the overall picture is summarised by Schmitt's (2008) overview (as cited in Eckerth & Tavakoli, 2012, p. 228) that incidental vocabulary learning from reading "is dependent on encountering a vast amount of textual input," does not contribute many new words, can improve orthographic and grammatical knowledge features of partially known words, leads to "word recognition gains that are more time-stable than gains in word recall," and therefore enhances the development of partial more than being complete vocabulary knowledge. Further details will be discussed in Section 2.5.3.

2.5 Vocabulary Knowledge and Reading Comprehension

Earlier research on L1 reading suggested that vocabulary knowledge and reading comprehension were closely related (Chall, 1987). As mentioned before, children's vocabulary size explains a significant proportion of reading comprehension performance and, conversely, having a good level of reading comprehension can make it easier to learn new words. Chall (1987) proposed that these two effects can be observed during different stages for young native speakers of English. When they start learning how to read, native speakers' vocabulary knowledge is an important determinant of their reading comprehension. As native speakers, they begin school with a vocabulary approaching approximately 5,000 word families, a target which is, according to Chall (1987), not difficult to reach. After learning how to read, the relationship changes. This means that for skilled readers, reading can become a source of vocabulary learning.

Stanovich (1986) pointed out in his seminal paper on the Matthew effect (i.e., that the rich reader will become richer and readers who are poor will stay poor), that early failure to acquire word and reading skills in L1 affects reading acquisition. He argued that if less able readers are continually denied chances to read actual texts, they will continue to fall further behind. Children with slower reading skills may find that, due to reading slowly, their reading skills are not built up in an exponential way and this will then lead to them having fewer opportunities to learn words from reading. This will result in negative attitudes towards reading and having smaller vocabularies than those with larger vocabularies, who can read better, which will lead to the latter reading more and more and so gaining larger vocabularies. An observation which supports this concept is the finding that an excellent middle school reader might read 10,000,000 words per year, whereas a poor reader may only read approximately 100,000 (Nagy & Anderson, 1984).

Researchers have recommended several models of the relationship between vocabulary knowledge and reading comprehension. Among them, Anderson and Freebody (1981) proposed a model called *the Knowledge Hypothesis,* which suggests that vocabulary knowledge is a key predictor and an important factor in reading comprehension. They claimed that the reader must have a good understanding of the words in the text before comprehending it.

Perfetti (1985; 2007) suggested that the lexical quality hypothesis confirms the significance of mastering word knowledge, i.e., a high level of word knowledge is the best way to achieve good reading comprehension. In order to show the relationship between vocabulary knowledge and reading comprehension more clearly, it would be useful to have separate

sections: the effect of vocabulary coverage on reading comprehension, the effect of vocabulary depth and size on reading comprehension, and the effect of reading comprehension on vocabulary learning. Further details about these three effects will be explained next.

2.5.1 The effect of vocabulary coverage on reading comprehension

As vocabulary knowledge is crucial for successful L2 reading comprehension, it is important to focus directly on lexical coverage, that is, the percentage of words that need to be known in a text if readers are to understand the text (Grabe, 2009; Koda, 2005). Several studies have established the relationship between vocabulary coverage and reading comprehension by defining the percentage of vocabulary needed by an adult second language learner to be able to read a text without unknown vocabulary being problematic (Laufer & Sim, 1985; Laufer, 1989; Hsueh-chao & Nation, 2000; Schmitt et al., 2011). These studies show that that there is no single coverage figure, although in general the larger the vocabulary size of an L2 learner, the better is their reading comprehension. For example, 98% coverage and higher may lead to a successful reading, while readers struggle to understand texts which contain more unknown words (Laufer, 1989; Hsueh-chao & Nation, 2000). This indicates that more vocabulary known is generally better, suggesting that there may be a number of different vocabulary thresholds for adequate text comprehension, depending on the level of comprehension that is required. One of these studies was conducted by Laufer and Sim (1985), who investigated the threshold level for comprehension of English texts for academic purposes in the First Certificate in English exam by utilising a comprehension test and interviews with learners. They found that minimum required scores of 65-70 on the comprehension test constituted the threshold. Following interviews and the comprehension test, they concluded that the most pressing need of second language learners was vocabulary, followed by subject background knowledge, and then syntactic patterns.

Subsequently, Laufer (1989) investigated how much vocabulary is necessary to obtain a score of 55% on a reading comprehension task. Participants were asked to underline unknown words in a text, and their responses were checked through a translation test which involved most of the infrequent words of the text. Then, a comparison was made between the underlining and the translation in order to reveal any discrepancies between these two tasks. Based on this, the percentage of vocabulary in the text each learner knew was calculated by converting the total number of words in the text minus the number of unknown words. It is found that 95% coverage was the point that can distinguish between learners who obtained 55% on the reading comprehension test versus those who did not. Using the 95% coverage, Laufer, based on Ostyn

and Godin's study (1985), clarified that about 5,000 words would constitute this vocabulary coverage.

Schmitt et al. (2011) stated that even though this is considered a good first attempt to specify the vocabulary requirements for adequate reading comprehension, many weaknesses exist. Laufer (1989) referred to the Ostyn and Godin's (1985) frequency counts, which are old Dutch frequency counts and used for second language learners of Dutch. Thus, it does not mean that they could be applied directly to English. Another important issue is that the comprehension criterion of 55% is very modest, and most readers might aim for a better understanding of texts.

The percentage of text coverage needed also depends on the kind of texts a reader wants to read. Hsueh-chao and Nation's study (2000) sought to investigate the required percentage of vocabulary coverage for adequate reading comprehension of fiction texts at 80%, 90%, 95% and 100% vocabulary coverage. It examined 66 participants on a pre-university course who were then split into four groups of 16-17 participants. A story, consisting of 673 words, was read by each group, at one of the above-mentioned vocabulary coverage levels. Participants were then asked to complete two measures of reading comprehension: a multiple-choice (MC) and a cued written recall (WR). The cued written recall (WR) comprises a set of questions about the story which aims to give L2 learners a chance to represent their understanding of the main ideas, followed with as many details as possible without time constraints. They found that at 80% vocabulary coverage (vocabulary known in the text), no learners achieved adequate comprehension on two of the relevant measures. With comprehension levels of 90%, only a few learners gained adequate comprehension, and most did not even reach adequate comprehension at 95%.

Simple regression analysis of the data represented showed that there was an expected relationship between the density of unknown vocabulary and the level of comprehension. It is suggested that the smallest percentage of vocabulary coverage to lead to comprehensible written input is 80% (i.e., one unknown word in five). The study also showed that 95% coverage will allow only 35% - 41% of the learners to gain reading with adequate comprehension. This means that for learners to obtain unassisted reading comprehension of a fiction text, approximately 98% - 99% coverage is required. Findings were based on only four coverage levels, to include a relatively small number of subjects in each coverage point (n=16 in each group).

Schmitt et al. (2011) directly investigated the relationship between vocabulary coverage and reading comprehension. They defined the specific percentage of known vocabulary required for adequate comprehension. In their study, 661 participants were selected from eight countries: Egypt, Spain, Turkey, Great Britain, China, Japan, Sweden, and Israel. Participants completed a vocabulary measure based on the words taken from two texts which they had to read before completing a reading comprehension test for each text. After reading the passage, learners were first asked to complete MC tests and then fill in the blanks in the graphic organiser (GO) completion test. The MC test included 14 MC questions. The GO, an information transfer test, is a more complicated test and requires increased cognitive processing. In the GO completion test, learners are also required to know the structured organisation of the text and logical relationships with the existing information, to find out where to put the information which belongs in the gaps. In Schmitt et al. (2011), participants filled in 16 blank spaces. Thus, the total of reading tasks included 30 items.

The results show a predictable relationship between the vocabulary known percentage and the level of reading comprehension. Thus, the percentage of known vocabulary requirement depends on the aims of comprehension. No evidence of a vocabulary "threshold" was found, where comprehension increased radically at a specific percentage of vocabulary knowledge. The results showed that 98% coverage is probably required, as 70% comprehension is desirable. However, even 90% coverage can lead to only 50% comprehension, and 100% coverage to only 75% comprehension. This is because there are other factors outside of vocabulary that can affect reading comprehension, such as language proficiency, the text itself, or the readers' motivation. However, a high vocabulary level is clearly required.

As most researchers agree that a large vocabulary is needed, it is important to find out which words learners use in the texts they produce. Learners' word selections, whether they use frequent or infrequent words, or whether the percentage of their using function words appropriately, would introduce helpful background knowledge about the learners' words resources. An important factor related to the learners' word selection is called "the measurement of word sophistication," which provides an indication of the percentage of low frequency words in a text (Milton, 2009). Laufer and Nation (1995) pointed out that learners normally tend to learn and use the most frequent words (in the first or second 1,000 General Service Word List) earliest. Lower- proficiency level learners can only use a small number of low frequency words in speech or writing. Advanced learners tend to know a larger number of infrequent words and are then more likely to use this resource in speech production or writing.

A question this raises here is whether spoken or written language will be more loaded with infrequent words. It seems that, as Milton (2009) suggested, written texts are more likely to be filled with low frequency words than spoken texts. A single written or spoken production may not be an accurate reflection of the learner's knowledge or ability because the vocabulary used in this text may be typical of one particular text genre only. Overall, studies which focus on the vocabulary that is deployed in oral or written language can provide important information about the levels of complexity and word sophistication in each. Such an analysis could be helpful in investigations how vocabulary knowledge is related to vocabulary use.

2.5.2 The effect of vocabulary depth and size on reading comprehension

Depth of word knowledge can also be acquired through extensive reading (Cobb, 2010). The following four related linguistic domains influence the learner's reading comprehension performance and vocabulary knowledge: phonology, morphology, semantics and syntax (Qian, 1998).

Phonology knowledge refers to the ability to successfully connect between graphemes (letters) and sounds and to be able to recognise the small segments of sound which make up a word. Studies of learners with learning difficulties in reading have emphasised the importance of phonemic knowledge in reading comprehension because these difficulties are caused by weakness in phonological coding (Adams, 1990). It has also been found that poor readers consistently under-perform in relation to the typical reader's level with regard to a phonological knowledge task (Hulme, Snowling, Caravolas & Carroll, 2005), indicating that there is a link between phonological knowledge and reading comprehension performance.

Morphology knowledge refers to the conscious ability to recognise the small units of meaning, e.g., suffixes, prefixes and roots, and to apply them in more complex structures. As an example, a reader can reveal morphological knowledge[1] by realising that the word *thoughtful* consists of the following two parts: the root *thought*, which is a noun, is combined with the suffix *-ful* in order to create an adjective (Carlisle, 2003). Several studies have consistently reported findings indicating that morphological knowledge is an important contributor to reading comprehension abilities (Kuo & Anderson, 2006; Kieffer & Lesaux, 2008). Furthermore, researchers have also hypothesised that the relationship between morphology and reading comprehension strengthens and increases over time (Kuo & Anderson, 2006; Kieffer & Lesaux, 2008).

[1] 'awareness' replaced with Nation's terminology 'knowledge'

Semantic knowledge refers to knowledge of the meaning of words and how words share semantic features with other words. The importance of word meaning for reading comprehension was also highlighted by Qian (2002) with a group of 217 ESL university students who showed that semantic depth of word knowledge, assessed by the Depth of Vocabulary Knowledge (DVK), explained more variance 13% than size 8% in reading comprehension. Therefore, there seems to be a positive relation between semantic knowledge and reading comprehension in ESL adult learners.

Syntactic knowledge refers to knowledge of language patterns. Such knowledge is necessary for learners to improve their vocabulary depth and comprehension (Wolter, 2006). A study conducted by Nergis (2013) with 45 undergraduate learners of English aimed to investigate whether depth of word knowledge, specifically the syntactic knowledge[2] or metacognitive awareness, was a more powerful predictor of academic reading comprehension. The result showed that syntactic knowledge significantly predicts reading comprehension in L2, however, depth of word knowledge was not a significant predictor of L2 academic reading comprehension. Low and Siegel (2005) explored the role of phonological processing and syntactic knowledge in reading comprehension of 884 English native speakers (NS) and 284 learners of English as a second language in Canada. They found that syntactic knowledge predicted and facilitated reading comprehension in both groups, and that the NS group surpassed the ESL group in knowledge of syntax. However, both groups were similar with respect to phonological processing and verbal working memory tests, which further predicted reading comprehension. Therefore, there is considerable evidence that there is a significant relationship between syntax and reading comprehension.

Based on the studies reviewed, it is clear that learners with greater knowledge of these domains are better at knowing words in depth, and each of these may contribute to successful comprehension performance. In contrast, many researchers have placed greater emphasis on the size of vocabulary (e.g., Liu & Nation, 1985; Laufer, 1992) than the depth of vocabulary knowledge (e.g., Paribakht & Wesche, 1997; Qian, 1998; 1999). Both vocabulary size and depth, however, are necessary for reading comprehension (Qian, 2002), which raises a question regarding which one of these dimensions contributes more to reading comprehension.

Rashidi and Khosravi (2010) explored to what extent depth and size of word knowledge among 38 Iranian university students would predict their L2 reading performance. All measures

[2] 'awareness' replaced with Nation's terminology 'knowledge'

were administered together within the same session. The Word-Associates Test (WAT) (Read, 1993) was conducted to measure depth of word knowledge, and the Vocabulary Levels Test (VLT) (Nation, 1983) was administered to test learners' vocabulary size, and the Reading Comprehension Test (RC) for measuring their reading performance. The results showed not only that there was a positive correlation between depth, size and reading comprehension, but also that depth of word knowledge contributed significantly to L2 learners' reading comprehension, over and above vocabulary size. After controlling the size of vocabulary knowledge, they found that depth of vocabulary knowledge explained more unique variance (69%) in reading comprehension performance than vocabulary size (55%). They further found that L2 learners who had stronger depth and size of vocabulary knowledge performed better on reading comprehension, which clearly indicates that depth and size are closely correlated, and both can enhance reading comprehension. In another study, Zhang and Yang (2016) examined the contribution of vocabulary depth of L2 learners of Chinese to their reading comprehension performance. The result revealed that vocabulary depth explained unique variance in reading comprehension over and above vocabulary size. More specifically, they found that in short passages which required inferencing knowledge, vocabulary depth was a stronger predictor than vocabulary size. These findings, indeed, revealed the importance role of depth of word knowledge in L2 reading comprehension.

Other researchers confirmed the significant role of vocabulary size in achieving successful reading comprehension (Hsueh-chao & Nation, 2000). Qian (2002) clarified that size and depth are closely related concepts because the greater the number of words that learners know, the better they know many of those words. The learner who, for instance, knows words such as *science, scientific and scientifically,* will increase the depth of that knowledge by recognising the common morpheme *science.* Additionally, with increased experience with a specific word, that word can be associated with other words in different contexts, thereby suggesting that depth can contribute to size as well (Haastrup & Henriksen, 2000). As suggested by Li and Kirby (2014), deeper knowledge of a word can lead to learners learning more words because they can make links between new and old lexical items. Then, as the number of words known increases, more extensive lexical networks which expand their word knowledge are also more likely to be developed. As an example, learners who learn the word *design*, for example, will increase the depth of that knowledge by learning the morpheme *sign*. This can facilitate recognising the relationship between words like *signature* and *assign* and support them learning the meanings of those words (Li & Kirby, 2014). Increasing the depth of knowledge and recognising the relationship between words are closely related to reading comprehension. Learners with a larger

vocabulary size will expand the depth of their vocabulary knowledge too, which will then also have a positive effect on their reading comprehension performance.

As shown in this section, there is evidence that size and depth of vocabulary knowledge are uniquely related to reading comprehension. Size of vocabulary knowledge researchers placed great emphasis on the importance of size of vocabulary in order to achieve success in reading comprehension (Hsueh-chao & Nation, 2000). Conversely, increasing the depth of knowledge for a particular word would make it possible to use it more flexibly so that its semantic features would be accessed more readily in multiple contexts (Anderson & Freebody, 1981). What the relative contribution of size and depth to reading comprehension is, therefore, remains vague, in part due to the way in which researchers operationalise these dimensions.

2.5.3 The effect of reading comprehension on vocabulary learning

In L1 acquisition research, it is clear that reading is one of the key ways of learning new words, and more reading can lead to an increased vocabulary (Strenberg, 1987; West & Stanovich, 1991). This was confirmed by Saragi, Nation and Meister (1978), who asked adult English native speakers to read a novel that contained 241 Russian slang words (nadsat words). Participants were not told to focus their attention on the nadsat words; instead, they were only informed that after they completed reading the book, they would take a comprehension task based on the novel. The participants did not have access to a dictionary that included these words. A few days later, participants were given a multiple-choice test with over 90 of the target words. The lowest score on the multiple-choice test was approximately 50% correct, and the average score was approximately 76% correct. These results suggest that many of the target words were learned.

Many early studies have examined whether L2 learners can acquire vocabulary through reading (Pitts, White & Krashen, 1989; Day, Omura & Hiramatsu, 1991; Horst et al., 1998). However, these early studies found a discouragingly low absorption rate of approximately one word being learned out of every twelve words examined. However, these studies had a number of methodological issues, including insufficiently controlled text difficulty, inaccuracy of measuring instruments, short texts, a small number of target words, and usage of a short reading passage. Studies that are more recent have attempted to overcome some of these issues by focusing on reading that is more extensive and applying new measuring instruments. These showed more gains from reading than previous studies recorded.

One of these earlier studies, Horst et al. (1998) asked L2 students to read a 109-page book (a simplified version of *The Mayor of Casterbridge*) over a period of ten days. To ensure that the participants read all 21,232 words of the text, the teacher read the entire text aloud in class while the participants followed along in their books. The participants were then examined using 45 MC items and a 13-item word-association test. The results indicated a gain of approximately one new word out every five words, and this learning persisted over a period of ten days.

Another study was conducted by Pigada and Schmitt (2006), who examined the learning of spelling, meaning and grammatical knowledge over a one-month extensive reading period. They revealed that knowledge of 65% of the target words was enhanced with respect to at least one of these aspects of word knowledge. Even through a small number of exposures, knowledge spelling strongly increased. Additionally, knowledge of meaning and grammatical aspects was enhanced, but to a lesser level. With respect to vocabulary acquisition, they found that approximately 1 of every 1.5 words examined was learned. Horst (2005) concluded that her subjects gained over half of the unknown words they encountered in their extensive reading. Brown et al. (2008) conducted a study to investigate to what extent English vocabulary can be acquired from three input modes: reading, reading while listening, and listening to stories. The study was conducted with 35 Japanese undergraduate students of English literature at a private university in Japan. Two test types were administered: the form and meaning recognition multiple-choice (MC) test, and the meaning-translation test. The two tests were taken at three sessions: immediately after the tasks, a week later, and three months later. The results showed that learning new words incidentally can take a place in all three modes, but at a limited rate. By controlling to the frequency of the occurrence factor, it was found that words that occurred more frequently in the text were better learned and retained. An important finding was that more enhanced incidental vocabulary learning from reading was shown in testing the word form and meaning recognition in a MC test, but far less with respect to producing the meaning in a translation test.

In a study which examined the vocabulary learning of 15 participants who read one graded reader, *A Little Princess*, Waring and Takaki (2003) changed the spelling of 25 words from the text, making them into non-words, to ensure that each item was unknown. Three test types were conducted in three subsequent test periods: a simple recognition test, a multiple-choice test, and a translation test. They found that a few of the new words can be learned and retained incidentally from reading. The subjects realised 10.6 out of 25 words' meanings immediately on a MC test. However, they were able to translate from English to Japanese only 4.6 out of 25 words correctly. Three months later, even though the recognition of meaning score dropped to

6.1 words, the translation score decreased much more (0.9 words). This reveals that learning new words from reading incidentally is subject to forgetting. Overall, their results clearly show that using more than one type of test can help in gaining a more balanced picture of learning, as each test reveals another fact about the kinds of learning that can occur (Waring & Nation, 2004). Hill and Laufer (2003) found that, based on findings of incidental learning studies, L2 learners would need to read about 8 million words to increase their vocabulary size by 2,000 words.

An important factor related to vocabulary size which strongly affects vocabulary learning from reading is the frequency of exposure. For example, Rott (1999) found that six repetitions result in better learning than two or four repetitions. Pigada and Schmitt (2006) found that there was no frequency effect where meaning acquisition was assured, however, after more than ten exposures, an encouraging rise in the learning rate was found. Nevertheless, even after more than 20 exposures, the participants still found that the meaning of some words eluded them. Further, the participants in Saragi et al. (1978) learned 93% of the words presented more than six times, but words shown fewer than six times were learned by only half of the participants. Horstet et al. (1998) also found that after eight or more exposures to the new words, there was a reasonable chance of these new words being learned. Waring and Takaki (2003) found, however, that for the participants to have approximately a 50% chance of recognising a word's form or its meaning, at least eight repetitions were needed. However, even if new words were repeated 15 to 18 times, there was less than a 10% chance that these were successfully translated three months later, and no words repeated fewer than five times were translated correctly. This clearly shows that although higher repetition rates are valuable, learning a word may depend on more than just frequency of exposure.

Overall, the vocabulary research confirms that L2 learners can acquire vocabulary through reading, but the rate of learning (and retention of that learning) is low (Horst et al., 1998; Waring & Takaki, 2003; Horst, 2005; Pigada & Schmitt, 2006; Brown et al., 2008). These studies, using a variety of test types, have also found that the frequency of word exposure is important to the learning of new words and leads to developments in multiple aspects of word knowledge. This conclusion is enhanced by the view of Waring and Takaki (2003) that reading does not help L2 learners to learn a large number of unknown words. However, it is very helpful in improving and enriching their known vocabulary.

2.6 Forgetting Curve

Another important aspect in L2 vocabulary retention and memory is investigating how well the newly words learned are retrieved. In 1885, Ebbinghaus (as cited in Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006) analysed the course of memory and knowledge retrieval which is known as the forgetting curve. The general feature of forgetting after learning a piece of information drops very fast, but that it then, after regular repeated exposure to the items, does so slowly. Ebbinghaus (1885; as cited in Cepeda et al., 2006) also investigated that knowledge retrieval could be dramatically developed through the "spacing effect," which refers to the correct spacing of practice sessions. Basic to this forgetting curve, Pimsleur (1967) created his learning model, named the Graduated Interval Recall Model, which means that learners should revise immediately, when they still have an opportunity to know the meaning of approximately 60% of a word. When the learners, at that moment, cannot recall the meaning of a lexical item, they will be presented with it at another time, which will make the knowledge level again reach 100%. This is supported by the fact that knowledge declines far less quickly after each revision and that learners can then increase the interval between the reviews.

Landauer and Bjork (1978) provided a memory technique, called the Expanding Rehearsal Strategy, which was based on the importance of testing any new words after a short delay. If the learners recall the given item correctly, the delay should be scientifically increased, however, if they are incorrect, the delay should be decreased. In short, learners who are tested at a time when they can recall a word will be better able to retrieve the given items. Additionally, as the learning progresses, this delay will be increased, suggesting that there will be adequate room left to introduce new items (Bjork, 1988). Although how well the new words are retained seems to be an important and more effective aspect in L2 vocabulary retention, it is not often examined. There is, therefore, a knowledge gap which needs to be filled with respect to the rate of forgetting that may accrue from the learning from context.

2.7 Depth of Processing Hypothesis

To know about the memory process that deals with explicit information, Craik and Lockhart (1972), in a seminal paper, proposed the "Depth of Processing Hypothesis". The basic idea is that the chance of storing new incoming information into long-term memory is determined not by the length of time but by the depth with which it was processed initially. Elaboration reflects the richness of processing inside one level. Craik and Lockhart (1972) further proposed that new information progresses through several stages of processing depth. As an example, the phonological word form features processing at a shallow level, followed by

its meaning at a deeper level. Craik and Lockhart (1972) pointed out that the nature of processing information can affect learners' retention; words for which users need to access the meaning require more elaborate processing and have a higher retention rate than words which are only processed phonologically. Craik and Lockhart's hypothesis, however, was criticised by Nelson (1977) and Baddeley (1978), for two main reasons, which were addressed as: it is difficult to determine levels of processing, and it is not easy to know which level is deeper than another. Craik and Tulving (1975) proposed that retention can be affected by the presence or absence of depth (a quality) of semantic processing and the richness (a quantity) of elaboration processing during which the information is encoded.

To simplify the depth of processing notion in order to employ it within the framework of L2 vocabulary acquisition, Laufer and Hulstijn (2001) introduced the Involvement Load Hypothesis, which claims that incidental L2 vocabulary learning and retention is based on the involvement load induced by different post-reading vocabulary tasks. They assume that a greater depth of processing of the new words learned will lead to better retention. Further details on this theory will be explained in the next section.

2.8 The Involvement Load Hypothesis

Laufer and Hulstijn (2001) claimed that learning and retention of L2 words is influenced by the involvement load induced by a task. In other words, the effectiveness of a vocabulary task is determined by the three components of a task-induced involvement load: need, search and evaluation. *Need* is the motivation dimension of involvement load and refers to the learner's motivation and requirements to complete a task. It can be at the moderate level, when the learner was pressurised by an external condition, such as by the teacher, or at a strong level if the learner is self-motivated. *Search* and *evaluation* are the cognitive dimensions of involvement. *Search* indicates the efforts learners make to find the meaning of a new L2 word or the efforts to find a particular concept represented by a new L2 word to complete the task. It can be present, e.g., if the learner tries to find the meaning of unknown L2 words, or absent, when no search effort is required because glosses are provided at the end of the text. *Evaluation* refers to a comparison process whereby a learner is required to compare target words and other words to find the most relevant words for a particular context. A moderated *Evaluation* level would be, for example, recognising differences between words, and a strong *Evaluation* level is present in tasks which require the learner to decide which words can be combined with the target words in a given text. A task can include any one or all three of the components of involvement: need, search, and evaluation, for each word (see Table 2.2).

Table 2.2 Three components of involvement (adapted from Kim, 2008)

Components	Need	Search	Evaluation
Feature	Motivational, non-cognitive dimension of involvement	Cognitive dimension of involvement	Cognitive dimension of involvement
Operationalisation	Need to know words for the task	Attempt to find the meaning of unknown words. Attempt to identify the appropriate L2 form for a particular concept	Specific meaning of a word with its other meanings
Categories	Absent (0) vs. present: moderate (1) or strong (2)	Absent (0) vs. present (1)	Absent (0) vs. present: moderate (1) or strong (2)

To test the effectiveness of using vocabulary tasks with different involvement loads, Laufer and Hulstijn (2001) reviewed previous research in the field of cognitive and motivational aspects of L2 vocabulary learning. In order to overcome the vagueness and the lack of actual mechanism in the depth of processing notion, as proposed by Craik and Lockhart (1972), Laufer and Hulstijn (2001) used all these general cognitive notions in a framework and formulated the ILH which they set out to test empirically. The hypothesis builds on different earlier studies which looked at the effectiveness of different tasks in enhancing word learning. Hulstijn (1992), for example, found that learners who were asked to choose the correct meaning of the new words from multiple choice retained words better than when they were provided with a word synonym in a reading task. Another relevant result was obtained by Hulstijn and Trompetter (1998), who found that using an L1-L2 dictionary to write the words in a composition was more effective than using an L2-L1 dictionary for comprehension purposes. This means that writing compositions may require a stronger *Evaluation* level than reading. The role of dictionaries in learning vocabulary from reading was also investigated by Cho and Krashen (1994). Their results showed that using the dictionary consistently and writing the new words in sentences as examples in a note lead to more vocabulary learning than reading only, which can be interpreted as evidence for the effect of levels of *Need* and *Evaluation* in a task on vocabulary learning. Although these studies did not test the Involvement Load Hypothesis directly, Laufer and Hulstijn (2001, p.12) clarify that their findings suggested that "the superior task required a deeper level of processing of the new words than the other task."

2.9 Empirical Investigations of the Involvement Load Hypothesis

Hulstijn and Laufer (2001) conducted their own study with two groups who were at a high level of proficiency in learning English as a Foreign Language (EFL): 79 in the Netherlands

and 128 in Israel. Both groups received three tasks which induced different involvement loads to test the potential learning and retention of 10ten target words. The first task was a reading comprehension task, including ten multiple-choice comprehension questions with marginal glosses, which means that it induced moderate *Need* but no *Search* or *Evaluation* were required. The second task was a reading comprehension text with fill-in-the-blank gaps which required moderate *Need* and moderate *Evaluation.* The third task consisted of writing a composition in which students were required to use the target words which induced moderate *Need,* no *Search,* and strong *Evaluation*. A delayed post-test was conducted with the two groups: two weeks later in the Netherlands and one week later in Israel.

The results revealed that, as expected, composition tasks which induced higher involvement loads led to highest retention, whereas lower retention was found in reading tasks with fill-in-the-blank gaps, and lowest in the reading multiple-choice comprehension questions. The results were also slightly different in the two experiments, as the ILH was fully supported by the Hebrew-English experiment and partially supported by the Dutch-English experiment. The results of the Dutch-English group showed that Task 3 produced higher retention than Tasks 1 and 2 but Task 2 did not significantly result in higher retention than Task 1. As mentioned by Hulstijn and Laufer (2001), some defects can be noticed in the study, which required further research to validate their findings. As an example, the researchers did not use a pre-test to assess the participants' familiarity with the target words, and prior knowledge was tested by a post-exposure questionnaire. Time-on-task was also not controlled. Another problem with the study was that, as pointed out by Martínez-Fernández (2008), the word type in Hulstijn and Laufer's (2001) study was not controlled; the target words used in their study consisted of expressions and different classes of words. It was argued that further research was needed to create tasks which can induce different involvement loads in order to find out how they affect the incidental vocabulary learning.

Keating (2008) also examined 79 beginning learners of Spanish on three vocabulary tasks that induced different involvement loads, starting with a reading comprehension task, then a reading comprehension task including the target words, followed by the highest involvement task, a sentence writing task. Learners' retention was tested immediately after the task and again two weeks later. Keating (2008) found that the sentence writing task, which induced a higher involvement load, led to better retention than the other two tasks. The results also confirmed the importance of the amount of time taken to complete the task and its effect on learners' performance. One of the studies which did not support ILH is the one by Folse (2006). Target words were practised in three different task conditions: one fill-in-the-blanks task, three fill-in-

the-blanks tasks, and a task which required learners to use the target words in a sentence. The results revealed that the three fill-in-the-blanks tasks (condition 2) led to better learning of new words than the other two conditions, which indicated that, as claimed by Folse (2006), word learning is affected by repeated exposure rather than depth of word processing (degree of involvement). The study's results, however, focused on testing the participants immediately, which may affect the results negatively, as it is possible that the effect of different involvement tasks may not appear immediately but at a later stage (Kim, 2011).

Kim (2008) conducted two experiments with L2 learners of English who were at two different proficiency levels. Experiment 1 contained three tasks which induced different levels of involvement, modelled on the original study by Hulstijn and Laufer (2001). Experiment 2 contained two tasks which induced the same level of involvement: writing a composition or new sentences using the target words. The results of Experiment 1 confirmed what Hulstijn and Laufer (2001) had found, where L2 learners who were assigned a higher involvement task, the composition group Task 3, performed better in learning the target words. The other two groups who were assigned Tasks 1 and 2 did not show any significant differences in learning the target words (in the immediate post-test). The results of the study (delayed post-test) fully supported the ILH; the task which induced a higher involvement load led to better retention, which indicated, as suggested by the author, the effect of different involvement tasks may not be shown straightaway but at a later stage. Kim (2008) also found that Experiment 2 resulted in similar scores in both tasks, which showed that there was no relation between the task which induced a particular involvement and L2 proficiency in learning L2 vocabulary.

One of the limitations suggested by Kim (2008) against the assumption made by Laufer and Hulstijn (2001) is that each component's level has the same contribution to the overall involvement index of the task. For example, a task which involves a moderate *need* (1), *search* (1), and no *evaluation* (0) has the same involvement index (=2) as a task which contains a moderate *need* (1), no *search*, and moderate *evaluation* (1). The author argues that these two tasks which have the same involvement index (2) but involve a different distribution of components might contribute to an overall involvement load differently. This means that they might not contribute to vocabulary learning equally. Kim (2008) suggests therefore to develop the ILH by testing tasks with different degrees (moderate and strong) of the same components in order to determine an accurate level of their effectiveness for vocabulary learning.

Another research to test the validity of the Involvement Load Hypothesis was carried out by Yaqubi et al. (2010). The study examined 60 EFL learners who were divided into three

groups. Group 1 received a reading comprehension task which contained multiple-choice items (with an involvement index of 3), Group 2 also completed a reading comprehension task with multiple-choice items (but with an involvement index of 2), and Group 3 completed a reading comprehension task and had to write one to three essays using the target words (with an involvement index of 3). The results did not support the ILH, as they found that the group who completed Task 2 (Group 2) with a lower involvement index obtained higher scores than the group who carried out Task 1 (Group 1), which induced a higher index. They also claimed that Group 3, who completed an output-oriented task obtained better results than Group 1, who carried out an input-oriented task, regardless of the involvement index; this contrasted with Laufer and Hulstijn's (2001) claim that with the same involvement load, these two types of tasks are equivalent.

Another contribution to the validity of the Involvement Load Hypothesis was conducted by Eckerth and Tavakoli (2012), who investigated the differential effects of word exposure frequency and elaboration of word processing among advanced L2 learners of English. The sample was 30 students at a UK university. The results revealed that exposure frequency and elaboration of word processing produce similar effects in initial word learning. Subsequently, elaborate processing of form-meaning resulted in more effects compared to the word exposure frequency in word retention. However, the researchers clarified that because the environment of study was inside a classroom and there was an immediate post-test which may have attracted learners' attention to the need to learn vocabulary, it is difficult to assume that L2 learners learn these words incidentally. Yet, it is worth mentioning that being tested inside the classroom does not necessarily mean that this would be under classroom conditions.

Eckerth and Tavakoli (2012), however, mention a number of limitations related to the ILH. The authors argue that the Involvement Load Hypothesis does not consider some factors involved in incidental vocabulary learning through reading. Firstly, we cannot be sure whether students read the glosses given in the task, or how long they took at glosses and target words. Secondly, variables such as the level of difficulty of a fill-in-the-gap task, how many contextual clues are included, or the ability of students to guess need to be taken into account. Thirdly, if students derive a meaning and retained such meaning over time, in this case, the measure is not on learning-retention, but on inferencing ability. Eckerth and Tavakoli (2012) suggest further investigations focusing on finding how students perceive and complete tasks and also considering other factors that may facilitate their vocabulary acquisition process.

Another study was conducted by Tang and Treffers-Daller (2016), who examined 185 EFL Chinese learners at a secondary school in China. This study was the first one which tested the ILH with low proficiency students in English. Six tasks with different involvement loads were designed by the researchers and were completed by six groups. The results fully supported the ILH in that tasks with higher involvement loads led to learning and retention of more words from reading. The results also revealed that *Evaluation* was the most important component of involvement to learn and retain new vocabulary, followed by *Need*, whereas the least important of the three was *Search.* This confirmed Laufer and Hulstijn's (2001) suggestion in that *Search* might be the least important component compared to the other two; Kim (2008) also showed that the task with a stronger *Evaluation* index can lead to processing a word with a greater involvement than the moderate *Evaluation* level and also the other two components. Tang and Treffers-Daller (2016) also found that learners who completed output-oriented tasks gained lower scores than learners who carried out input-oriented tasks with the same involvement index. This revealed that, contrary to the findings of Yaqubi *et al*. (2010), learners who took an output-oriented task gained better results than the group who carried out input-oriented task, regardless of their involvement index.

In conclusion, the ILH provides an effective theoretical framework in the field of L2 vocabulary acquisition (Laufer, 2012), and the cogency of this theory was found by testing five important factors: task design, task type, time-on-task, the text itself and word ratios (Huang, Eslami, & Willson, 2012). However, as argued by different authors (Laufer & Hulstijn, 2001; Kim, 2011) more empirical research designed to examine the claims made by the ILH is greatly needed.

2.10 Issues of Reading Comprehension Performance for L2 Learners

Various studies have shown a strong relationship between vocabulary and reading comprehension. However, the causal relationship is very complex. For example, Mezynski (1983) suggests that the access view of this relationship views vocabulary knowledge as being a major causative factor in comprehension, provided that the vocabulary can be easily accessed. This access can be affected by several factors, including fluency of lexical access and speed of word recognition. It seems that L2 learners need to build bilingual experiences to engage in translating difficult items. A study conducted by Kaushanskaya and Marian (2009) investigated whether bilingualism facilitates learning of novel words in adults with different languages, monolingual English speakers, early English-Spanish bilinguals, and early English-Mandarin bilinguals. The results showed that both bilingual groups outperformed the monolingual group,

which suggests a general bilingual advantage for L2 adults' performance. As indicated by Kaushanskaya and Marian (2009), early experience with two languages in life may facilitate the capacity to learn new words in adulthood. The bilingual advantage in learning new words may result from bilinguals' increased memory storage ability compared to the monolinguals. This was originally proposed by Papagno and Vallar (1995), who pointed out that the L2 learners' advantage for word learning is basically derived from their higher working memory ability. This was confirmed by previous research (Service, 1992; Gupta, 2003) by indicating a relationship between word-learning and working memory performance.

Other researchers viewed that reading abilities in L1 seem to be radically different from reading in L2, as in L1 reading, translating the written sequence of letters is not needed because it is the mother tongue for reading. For example, Carrell and Grabe (2002) noted a number of differences while reading in a first language compared to a second language, including language-related issues, experiential, and socio-cultural differences. Hsueh-chao and Nation (2000) pointed out that the relationship between vocabulary and comprehension is even more complicated for L2 learners. These complications can result from the common situation of being a L2 learner and beginning to read with limited vocabulary knowledge in the second language. A different writing structure in the second language compared to the first language can also be problematic. Recent research investigating the effect of script on language selection suggests that L2 learners whose two languages have different scripts (e.g., Chinese-English or Japanese-English learners) are not able to use script adequately. They are as a cue, allowing early selecting of one language (Hoshino & Kroll, 2008).

On the other hand, L2 learners with two languages which share a language script, such as Spanish-English learners, do not have a salient cue accessible to them during visually processing either language. When one language has less useful information sources available, same-script learners may need further control in order to help them restrict cross-language competition (Linck, Hoshino & Kroll, 2008). Accordingly, it is clear that same-script learners show greater control when compared to different-script learners. This indicates that a long distance between two languages can pose problems for L2 learners. As in the case of the current study, different scripts between English and Arabic languages may cause some difficulty for L1 Arabic English learners in acquiring English script. A more detailed discussion will be clarified next.

2.10.1 Issues of reading comprehension performance for L1 Arabic learners of English

To understand how L1 literacy affects L1 Arabic learners' ability to read in English, it is important to examine L1 Arabic orthography and Arabic literacy education. In the Arabic language, there are two main forms: Modern Standard Arabic (MSA), which is the formal form of Arabic, and a colloquial dialect. MSA is quite different from colloquial Arabic in terms of vocabulary and in some features of grammar and phonology. For all written and reading texts, L1 Arabic children learn to read and to write in MSA. However, they tend to speak in colloquial Arabic (Saiegh-Haddad, 2003).

Arabic is written from right to left, and the letters must be joined in a cursive form. MSA uses an alphabetic orthography that includes 28 letters. These letters are initially classified as consonants, however, they also include three letters that match long-vowel phonemes. There are also another three diacritics, which are short vowels that are written above or below the consonant letters (Bauer, 1996). In MSA, all words are based on a root morpheme, which generally consists of three or four consonants (e.g., *k-t-b* is the general morpheme for the concept to *write*). The root morpheme must be manipulated in a fixed pattern, including the diacritics, to generate a word (e.g., *kataba* wrote, *yaktub* writes, *kitaab* book). Even after adding letters and diacritics to the root morpheme, the root morpheme consonants continue to be present in similar order (Shimron, 1999). Therefore, when reading in Arabic without diacritics, a reader must include extra lexical knowledge, such as syntactic structure and morphological knowledge (Abo-Rabia, 2002).

As spelling is closely related to word recognition skills, lack of spelling knowledge can lead to deficiencies in word recognition and in reading comprehension. From a component skills approach, reading comprehension is the production of word recognition, which includes spelling representations and listening comprehension (Grabe & Stoller, 2002; Ehri, 2005). Ryan and Meara (1991) used a same-different matching judgment task with three groups: L1 Arabic learners of English, non-Arabic speakers ESL who matched their proficiency level and an English native group. The task included a one-second presentation of long words containing ten letters such as: *department*, *experiment* or *revolution*, then followed with a two-second empty screen, followed by the presentation of the same words but with incorrect spelling, e.g., with a vowel missing in one of four places. The vowel missing in some of the words was in the second letter place: *dpartment*; some in the fourth letter place: *expriment*; others in the sixth letter place: *revoltion*; or in the eighth letter place: *photogrph*. The ESL subjects were then

required to decide whether the second word presentation was a correct or incorrect spelling when missing a vowel letter.

The results showed that L1 Arabic learners of English made significantly more errors than either the non-Arabic ESL or native English speakers group. They also took significantly longer than the other groups to reach a judgment. Ryan and Meara (1991) claim that L1 Arabic literacy can affect Arabic ESL learners' performance in processing some of the vowel instruction encoded in the graphic display. They showed that since media materials, such as magazines and newspapers, do not insert short vowels, L1 Arabic readers learn to focus more on the consonant structures of the Arabic words, particularly as the consonant forms represent the root morpheme, and associated semantic information. They also clarified that L1 Arabic learners of English tend to transfer their L1 recognition pattern for words to English, relying more on consonant graphemes in word spelling than on vowel graphemes. Therefore, L1 Arabic learners of English tend to spell consonants more correctly than vowels in English, especially in unstressed syllables (e.g., *done, gone, phone*).

Abo-Rabia (2002) argued that the Arabic word recognition process is based on realising the root morpheme and employing the semantic and syntactic context to deduce the missing information about short vowels. Fender (2003) showed that L1 Arabic learners of English were slower than a group of Japanese ESL learners, who matched their proficiency level, in a lexical word recognition task. The results indicate that Arabic learners have slower and less well-organised word recognition skills, which indicates weak orthographic knowledge or spelling patterns. Studies conducted by Fender (2003), and Ryan and Meara (1991), therefore suggest that L1 Arabic learners of English have more conflicts with word recognition than their ESL peers, even when matched for proficiency. This might be due to the L1 transfer of the word recognition process to read the unclear unvowelised scripts, which may create difficulties for those concerned when realising and acquiring English word forms.

Part of this difficulty may also result from the more difficult script in L1 Arabic literacy which has missing vowel information, but which nevertheless has fairly reliable grapheme-phoneme mappings. There are, however, some morpheme-grapheme spelling variables in English, even for consonants, as the phoneme /k/ being implied as c as in *picnic*, ck as in *stuck*, and the grapheme *gh* being uttered as /f/ as in *laugh*, or as in the more complicated part which has no grapheme-morpheme mapping, as in *caught* (Fender, 2008). Thus, L1 Arabic learners of English may also struggle in acquiring English spelling scripts that diverge from steady phoneme-grapheme mappings.

2.11 Testing Vocabulary Knowledge

As vocabulary knowledge is a multidimensional construct (Nation, 2001), different measurements are needed to measure these dimensions. There are many tests that are reported to be valid and reliable, at least in the case of the goal for which they were created. In this part of the literature review, first, it is important to discuss some characteristics that play a central role in selecting a valid and reliable test. Then, the focus will be on discussing reading assessments and some of the offline and online tests of vocabulary knowledge and digit span tests that are directly relevant to the purpose of the current research and to the tests conducted. Reviewing these tests will help identify the measures that can be used to investigate the vocabulary knowledge of the students who participated in the current study.

2.11.1 Reliability and validity of a test

The selection of an assessment test is a crucial issue because the accuracy of the outcomes of a study is largely dependent on the test chosen for assessment. Thus, the assessment test selected should have two important features: *reliability* and *validity*. A test is considered to be reliable if it satisfies three key features:

Reliability of a test means being able to conduct a task or measure a task with consistency and accuracy. Test is considered to be reliable if the test is able to replicate or at least fetch similar results when the test is applied at different instances where the input conditions of the tests have not been changed (Milton, 2009).

Parallel form means the test can be conducted in parallel with different versions of the assessment tool with a group of students to evaluate its consistency across different versions. If the outcomes of the tests are in correlation with each other, the test then is reliable (Burnett, 1974).

Inter-rater reliability means the degree to which the different assessors agree with the scores obtained from the assessment tool. The inter-rater reliability is important because there could be differences in the way the assessors may interpret the scores obtained. Thus, a reliable assessment tool would be successful in reducing the disagreement of the assessors (Phelan & Wren, 2006).

Validity is another essential feature of a good test. An assessment tool can be valid if the tool is able to accurately measure what it is supposed to measure. The result of an assessment tool is not valid if it does not capture the required aspects (Milton, 2009). For example, an assessment test designed for particular language learners or age groups may produce invalid

results if used on a different set of language learners or age group. In such cases, the results of the test are invalid, even though the reliability of the assessment tool is not compromised (Fitzpatrick, 2007). There are various types of validity. Some of the important ones are as follows:

Construct validity is perhaps the most essential type of validity, and it can be subsumed under all the other types (Messick, 1989). It refers to the extent to which an assessment test measures the concept or the construct of the study that it is expected to measure. The assessor should ensure that the test measures what it is aimed to measure, excluding irrelevant variables. The construct validity of an assessment tool should be examined by experts or even the target learners to ensure that the aspects the tool measures are related to the goal of the assessment study. Although construct validity is considered to be the key type of validity, it is the most challenging type to address. For example, testing the construct validity of the tests used in the current project seems to be quite a challenging issue because the tests have to be inferred from the language produced by learners, which might be affected by other factors or conditions beyond the assessor's control. One main factor that may, to an extent, affect the reliability and validity of language testing is the test takers themselves. Nation (2007) clarifies that the validity of any test is based on the language learners' seriousness about showing their language knowledge accurately.

Content validity is defined as "the extent to which the test incorporates a representative sample of the entire domain being investigated" (Hughes, 1989, p. 22). Milton (2009) proposed that if the test has a good content validity, it means that it includes the appropriate content which covers the construct that it is supposed to measure. For example, the sample passage or text or words used in an assessment tool should be chosen in such a way that it is representative of the level at which the test is being conducted.

Face validity refers to the extent the test is expected to measure as seen by the test takers. For instance, the test takers may wonder about the simplicity of some reading comprehension tests which are based on careful selection data to provide a simple but accurate measure. They may not necessarily find the construct and content of the test good enough to measure their reading abilities. The assessors, therefore, need to consider the face validity of the test before selecting it for their target test takers (Gyllstad, 2007).

Thus, the *reliability* and *validity* are two crucial aspects during the process of selection of an assessment test. However, even if a valid and reliable test is conducted, it may not always provide identical scores, especially when a language is tested. In the light of the objective

testing in the current study, even though an attempt has been made to address the variables that may impact the participants' scores, there may be less control over other factors, such as their willingness or anxiety when performing the task. As pointed out by Nation (2007), some of the factors that might influence learners' performance are their attitudes and individual differences, which are difficult to control. Nevertheless, it is important to understand that the validity and reliability of a test should not be impacted by a few minor errors that learners may produce.

2.11.2 Reading assessment

Reading assessment is quite difficult because it is an unobserved process and only the product of comprehension performance can be observed (Fletcher, 2006). Brown (2004) pointed out that reading skills (micro-skills and macro-skills) and strategies (skimming and scanning) should be taken into account while measuring the learner's ability to make significant contributions to reading. Knowing which skills are assessed by a particular comprehension test is essential for both researchers and teachers. Magliano, Millis, Ozuru, and McNamara (2007) argue that it is essential to know the appropriateness of a reading assessment tool for assessing a type of learner or a group of learners. They propose a framework that enables choosing the right reading assessment tool for a given study. The assessment tool selected can be effective when factors such as the purpose of the assessment, the aspects of the reading comprehension that the tool is designed to assess, the content used in the tests, and the target examinees and their ability levels are considered before selecting a tool. Therefore, the reading assessment tool selected can be reliable and valid for a study if it is appropriate for the conditions of the assessment study.

Reading tests should measure the construct of reading, which means the test takers' reading ability. As Messick (1996) clarified, scores on reading tests that are designed to measure other abilities, unrelated to the construct of reading, may show construct-irrelevant variance. Construct irrelevant variance is considered as extraneous uncontrolled variables that influence assessment outcomes. The construct irrelevant variance may be present due to factors such as poor examination questions, unscientific formulation of questions, and irrelevant questions or measures (Haladyna, & Downing, 2004). For instance, the ability to use a mouse in a computer-assisted version of a reading test may have an effect on test takers' performance. Therefore, this task does not only assess the participants' reading skills, but also their computing skills, which does not belong to the construct the task is assumed to assess. Thus, it is important to use a valid assessment of reading ability.

Irrelevant variables, such as a test taker's experience with taking tests or their ability to guess the correct answer may lead to construct-irrelevant variance in the data, which adversely affect the reliability and validity of a test. Construct-irrelevant variance is considered as an error variance which particularly affects the construct validity of the tests because it reflects on the test taker's actual performance. By this way, the outcome may partly measure this irrelevant construct and may include more false negative data because the examinee might not be able to give correct answers, and this would impact their overall assessment (Haladyna & Downing, 2004; Leighton & Gokiert, 2005).

In the light of reading comprehension assessment, several skills can contribute to reading ability, e.g., decoding, language proficiency, and working memory. However, the main aim of reading is comprehension. Therefore, the skills that are most relevant to the construct of the task that is being assessed should be taken into consideration in designing or selecting reading tests, and other variables such as working memory should be controlled for.

Furthermore, it is necessary to know how representative a reading test is of the leaner's skills, such as if a learner who shows poor reading comprehension on a particular test would also perform poorly on another reading test. Comparing the reading comprehension test to different reading tests, therefore, is another essential element in order to check its validity, and then it may help to obtain a learner's actual reading comprehension performance (Keenan, Betjemann & Olson, 2008). Given that the successful comprehension of the text requires accurate interpretation of the words in the text, it is important to focus on vocabulary knowledge in order to gain a better understanding of reading comprehension performance.

Although several tests exist which set out to assess reading comprehension, those tests are limited in some ways and are therefore less suitable. For example, Paulson and Henry (2002) supposed that the Degrees of Reading Power (DRP) designed to measure students' overall ability to comprehend, is an inaccurate and insufficient measure of a process as complex as reading. The York Assessment of Reading for Comprehension (YARC), published by Stothard, Snowling, Clarke, Barmby, and Hulme (2010) is based on the SVR model (Gough & Tunmer, 1986), in which reading comprehension consists of two skills: decoding and linguistic comprehension. The YARC Test seems to be an appropriate tool with which to determine reading comprehension for the following reason: as clarified by Martin (2011), it clearly distinguishes between the literal knowledge required or information already involved in the passage and those questions that necessitate inferencing for successful answers. In particular, the YARC test provides a clear introduction and a helpful description of how to use the test.

The YARC developers, further, administered studies and confirmed that it has acceptable reliability and validity. This will be further discussed in Chapter 3 (Section 3.3.5).

2.11.3 Vocabulary size measurements

Learners' vocabulary size is often considered the most important dimension of vocabulary knowledge and therefore this dimension is measured in many vocabulary tests (Nguyen & Nation, 2011). A number of tests that measure vocabulary size were designed as a diagnostic tool for teachers to observe learners' vocabulary growth for better classroom planning (Nation, 2001). However, vocabulary size tests now have come to be widely used amongst researchers to estimate L2 learners' vocabulary size for research purposes (Shiotsu & Weir, 2007; Milton, 2009). There are two main methods to estimate learners' vocabulary size. The first technique is by testing learners on a sample of words taken from a dictionary (Nation, 1993). The total score on the test represents the total number of words in the dictionary from which words in the test are chosen, e.g., if the sample consists of one in every 100 words in the dictionary, then the test taker's score on the test would be multiplied by 100 to estimate the overall vocabulary size. The second technique is based on corpus data from which a frequency list of words is extracted. The list may either be a general list, such as the *General Service List* (GSL), originally developed by West (1953), the *New General Service List* (Browne, Culligan, & Phillips, 2013) or a specialised one, such as *The Academic Word List* (AWL) (Coxhead, 2000). Vocabulary size tests which are based on frequency lists are designed on the same idea as the dictionary-based technique. This means that if a test taker knows a proportion of the words from a particular band that were sampled from all words in that frequency band, then it is expected that he/she will know a corresponding proportion of all the items in that band. Vocabulary size tests for L2 learners focus on a smaller range of words than for native speakers, as the lowest frequency words are much less likely to be known (Read, 2007).

The two most widely utilised lists of English words are the GSL and the AWL. The GSL includes roughly 2,000 word families. Word families (headwords) consist of all inflected and derived forms. The word *be*, for example, is high on the list and represented in all forms of the word, e.g., *am*, *is*, *are*, *was*, *were*, *being*, and *been*. The words in the GSL were selected on the basis of both frequency and importance. The GSL is divided into two parts: the first 1,000 most frequent words and the second 1,000 most frequent words (Milton, 2009). The AWL was created on the basis of frequency and range of words across the university classifications of humanities, science, commerce and law. This list includes 570 headwords which are not in the GSL but which occur in a wide range of academic texts (Nation, 2004). If the AWL is combined

with the most frequent 2,000 words from the GSL, it provides coverage of 86.1% of the words in a wide range of academic texts, as shown in Table 2.3

Table 2.3 Coverage of academic texts provided by the GSL and AWL

GSL 1 k	71.4%
GSL 2 k	4.7%
AWL	10.0%
Total	86.1%

While this coverage (86.1%) would help a learner to recognise most of the words in an academic text, it is still quite low compared with the coverage figures that are required for adequate comprehension (i.e., 95% and 98%). Milton (2009) suggested the following two reasons behind this: 1) since some of the texts demand more vocabulary knowledge, it would be difficult to understand the depth of many academic ideas without having a very large number of words to explain these ideas; and 2) specific academic departments might require their technical words, and thus learners would also need to understand these words in addition to the general and academic word lists in order to understand specific texts.

The British National Corpus (BNC) is also one of the most popular corpus of British English words. It contains 100,000,000 running words of English, 10% of which are drawn from spoken sources, with 90% from written language. The first two lists of 1,000 word families were created using a particularly designed 10 million token corpus. In comparing these three lists, it is noted that their major differences are the way in which vocabulary is categorised. Nation (2004) therefore asserted that it may be best for ESL learners, most notably beginners, to use materials based on the BNC list because it has slightly better coverage. As the GSL, an old list, plus the AWL, a list with a narrow focus, are in the three one thousand levels of BNC.

In the following section, two of the commonly used receptive vocabulary size tests will be discussed with reference to their proposed function and potential shortcomings.

2.11.3.1 Nation's Vocabulary Levels Test (VLT)

One of the well-known tests is the Vocabulary Level Test (VLT), first published by Nation (1983; 1990) and further developed by Schmitt, Schmitt and Clapham (2001), and is widely used by teachers and researchers in developing suitable vocabulary teaching and learning programmes for students. It is a receptive test which requires matching words with their synonyms or definitions or L1 translation. The VLT assesses test takers' knowledge at four

frequency levels of English word families: 2,000, 3,000, 5,000 and 10,000. It also includes academic vocabulary based on *The Academic Word List* (Coxhead, 2000), which is not based on frequency levels, but might be helpful for teachers in academic contexts. Findings from Read's (1988) study reveal that the test is reliable and participants' performance on the various frequency levels tends to fall into an implicational scale, which means the knowledge of the low frequency words indicates the knowledge of the most frequent ones. Furthermore, the VLT test has been successfully used for research purposes and found to be an informative measure (e.g., Cobb, 1997; Laufer & Paribakht, 1998).

The VLT was originally designed as a diagnostic tool, and, as stated by Nation (2001), its main goal is to allow teachers to explore quickly the frequency words, high or low, that learners know. Hence, one of the drawbacks of the VLT is that a gap was left between the fifth 1,000-word level and the tenth 1,000-word level, i.e., if learners achieved knowledge of the words in the 10k band, this predicts their knowledge of the words in the sixth, seventh, eighth and ninth 1,000-word levels. In other words, learners' knowledge of words from the tenth word level indicates their vocabulary size between 5,000 and 10,000 words. However, it does not provide information about students' knowledge of the intermediate levels, and the assumption that any words from the intermediate levels are known if words from the 10k band are known may be incorrect. Therefore, the VLT could be criticised for overestimating the word knowledge of L2 learners from intermediate levels.

2.11.3.2 Vocabulary size test (VST)

Nation and Beglar (2007) published another receptive vocabulary size test (VST) with a multiple-choice format, where each target word is presented in a short non-defining context with three distractors and one correct answer, as shown in the following sample from the fifth 1,000-word level:

1. time: They have a lot of time.
a money
b food
c hours
d friend

The VST is based on fourteen frequency layers of 1,000 word families in the British National Corpus (BNC). As stated by Beglar (2010), a word family can be an appropriate unit for a receptive vocabulary test because L2 learners beyond a beginning proficiency level have

some ability to control word building members and to classify both form and meaning based on relationships between regular affixes of a word family such as *produce, producing, producer*. Nation (2004) asserted that it may be best for ESL learners to use materials based on BNC lists because it has slightly better coverage than tests based on GSL and AWL lists. The VST test comprises 140 items with 14 levels of difficulty, becoming incrementally higher, with every level made up of ten items. Each item in the test represents 100 word families and there are ten items from each 1,000-word frequency level, that is, the test measures knowledge of up to 14,000 words. It therefore includes a wider range of frequency bands than the VLT, and words from all frequency levels up to the 14k level.

Nation (2012) noted that when assessing for receptive knowledge, it is better to have items that encourage retrieval knowledge and that allow for partial knowledge to be utilised to obtain the correct answer. Hence, the receptive vocabulary size test has a multiple-choice format that presents each target word within a short non-defining sentence, so that, as stated by Nation (2012), learners can have the orienting effect of the context, as well as the cuing effect of the options. Beglar (2010) gave five reasons behind choosing the multiple-choice format for the VST. Firstly, this format was selected to allow the content to be presented efficiently; second, to allow the test to be suitable for a variety of learners who have different background languages, as many learners should be familiar with the multiple-choice format; third, to control the degree of difficulty of each item; fourth, to help make the scoring as effective and reliable as possible; and finally, to encourage learners to demonstrate their knowledge for each item (Beglar, 2010). Furthermore, it was argued that using multiple-choice items can help different learners to respond to the test items in a somewhat similar way to ensure they use their test taking skills seriously (Nation, 2012). The VST will be further discussed in Chapter 3 (Section 3.3.4).

2.11.4 Vocabulary depth measurements

Compared to the efforts to improve various tests of vocabulary size, there has been rather less development in measuring depth of word knowledge, that is, how well words are known. The vocabulary size test typically measures students' ability to link the written form of a word to its meaning, whereas, as mentioned in Section 2.4.2.3, there is much more to know about a word: its pronunciation and spelling, morphological information, syntax, its frequency, and so on (Read, 2007). Read (2000) also stated that vocabulary is often shown in a range of contexts with different meanings, so that the word knowledge should be examined in specific contexts. Hence, it is proposed that a more fitting design to measure vocabulary knowledge would be

putting words into the correct context (Kudo, 2001). In the following, the focus will be on discussing the two widely used tests relevant to depth of vocabulary knowledge, they will be presented with reference to their purpose, use, and potential limitations.

2.11.4.1 The Word Associates Test (WAT)

One type of vocabulary depth of knowledge test is the Word Associates Test (WAT) (Read, 1993); it assesses test takers' knowledge of synonymy, polysemy, and collocation. The test is based on the concept of word association, consisting of a target word and six or eight other words; half of these are semantically and collocationally associated with the target word, while the remaining words are not associated. Thus, for example, if the target word is *beautiful* the eight answer options could be: *enjoyable*, *expensive*, *free*, *loud*, *education*, *face*, *music*, and *weather*. The test taker is required to choose four words that either represent a synonym or are used in association with the target word. In the example of *beautiful*, the synonym would be *enjoyable*, and the collocational associates would be *face, music* and *weather.* An added difficulty of the test is that the number of synonyms and collocational associates differ per test item. The test provides a chance to assess key dimensions of vocabulary knowledge that go beyond knowledge of words in isolation.

A modified version of the word associates test (Read, 1993; 1998) is taken by Qian (1999; 2002), named the Depth of Vocabulary Knowledge measure (DVK). Qian and Schedl (2004) claim that the DVK is a valid test to measure these three aspects of depth of word knowledge: synonym, polysemy, and collocation, as the examinees would be more likely to recognise the importance of multiple meanings of a word in filling in the DVK, as well as they have to understand the significance of rehearsed language use in order to deal with the collocation part of the DVK. Qian and Schedl (2004) pointed out that the DVK adopted from the WAT can be a feasible alternative to traditional multiple-choice items as vocabulary knowledge measures in testing English as a Foreign Language (TOEFL).

Some researchers, however (e.g., Bogaards, 2000; Greidanus, Beks, & Wakely, 2005), have also adopted Read's test and found the test to be very challenging even for advanced foreign language learners at university level. Further, it is unclear whether the word association test actually measures the learners' vocabulary depth in a second language or size. According to Milton (2009), the collocation of some words, such as *story* or *news*, might be very common in many languages, so that L2 learners understand the broad meaning of these words which could be derived from their L1. Therefore, the WAT might be more appropriate for studies assessing antonyms, synonyms or collocations of target words, and unsuitable to be

administered in research measuring other aspects of word knowledge, e.g., using the word in a sentence.

2.11.4.2 The Vocabulary Knowledge Scale (VKS)

Another measure of vocabulary depth is the Vocabulary Knowledge Scale (VKS) (Wesche & Paribakht, 1996), which was further developed by Brown (2008). This is a word knowledge test that asks learners whether or not they know a word, and also how well they know this word. The original VKS consists of a five-point scale ranging from complete unfamiliarity through to word recognition, to the ability to produce the word in a correct sentence with a grammatical and semantic accuracy, as shown in Table 2.4.

Table 2.4 Wesche and Paribakht's VKS (1996)

1. I don't remember having seen this word before.
2. I have seen this word before but I don't know what it means.
3. I have seen this word before and I think it means _____ (synonym or translation).
4. I know this word. It means _____ (synonym or translation).
5. I can use this word in a sentence: _____.

The VKS test measures depth of vocabulary knowledge and examines the learners' control of receptive and productive aspects of knowledge of the target words. Wolter (2005), however, identifies many criticisms which suggest the result of the VKS may be less useful than one might have hoped. As pointed out by Wolter (2005), the test seems insensitive to other aspects of word knowledge, such as the variety of word meanings which cannot be tested with these methods. It also seems that item 5 on the scale is seldom used and scores of 3 and 4 are also quite rare. The advantage of the test is, however, that it taps into different aspects of word knowledge and is not too complex even for learners with relatively low proficiency levels. The VKS will be further discussed in Chapter 3 (Section 3.3.8).

2.11.5 Online measurements

L2 vocabulary acquisition researchers have become increasingly interested in how lexical knowledge is organised and processed in the L2 mental lexicon (Jiang, 2000; 2002; 2004). The processes involved in word recognition can be investigated by conducting online tests, e.g., lexical decision and priming tests.

A lexical decision task (LDT) is commonly used to analyse lexical access and visual word recognition. With this task, a participant has to decide whether a string of letters presented visually is a word or not. Research using the LDT has identified that participants making lexical decisions recognise high-frequency words more swiftly than low-frequency words (Scarborough, Cortese, & Scarborough, 1977), which Gardner, Rothkopf, Lapan and Lafferty (1987), identify as the word frequency effect. Gardner et al. (1987) found that word frequency represents the familiarity of the participant with the word's meaning in the context of its most frequent occurrence. They asked three groups of participants (engineers, nurses and students) to make lexical decisions in relation to 720 words (360 words and 360 non-words) from six different categories e.g., engineering, medical, etc. Results showed that nurses recognised medical words more quickly, while engineers responded more swiftly to engineering words. The three groups responded equally to a standard neutral words list. This would represent an increase in the frequency of word usage, as a result of which it seems that the participant's lexical decision is affected by their experience with the word, so that greater experience results in a faster response.

Priming measurements, by contrast, allow the researcher to investigate the implicit relationship between the formal-lexical or semantic-lexical prime and the target. According to McRae, Hare, Elman, and Feretti (2005), priming research can provide data on aspects of lexical processing either in reading or listening, avoiding the confounds of additional restrictions that appear in normal reading to connected text or listening to connected speech. Both LDT and priming tasks allow the researcher to extract data directly relevant to the knowledge of the words, minimising chances for intentional and conscious control. They can also tap the accuracy, error rates and reaction times (RTs) which provide ways of measuring the levels of accuracy of identifying words and production (Elgort, 2011). The lexical decision task (LDT) and Semantic Priming Task (SPT), based on Forster and Davis (1984), are incorporated into the design of the present experiment and will be further discussed in Chapter 3 (Section 3.3.7), and Chapter 4 (Section 4.3.7).

2.11.6 Assessment of working memory

Working memory capacity has been shown to be associated with foreign language learning (Kormos & Safar, 2008) in that L1 and L2 learners with a strong working memory often obtain higher scores on language proficiency tests (Olsthoorn, Andringa, & Hulstijn, 2014). As the current study concerns L2 vocabulary learning and retention, it is important to understand the

role of working memory (WM) and how to control for this factor, particularly for Study 2, where both groups were L1 Arabic learners of English.

Earlier research proposed two theoretical components of the human memory system (Atkinson & Shiffrin, 1968): long-term and short-term memory. The conceptual difference between these two components is derived from the time period for which new information is stored. Long-term memory (LTM) refers to a system of unlimited and permanent storage of information accumulated over time, whereas short-term memory (STM) is the system which is responsible for storage over a short time period. Based on this model, cognitive processes such as rehearsal can transfer information from short-term to long-term memory (Atkinson & Shiffrin, 1968).

Baddeley and Hitch (1974) proposed the notion of Working Memory (WM), a limited capacity system which stores information temporarily, and which is distinct from STM. WM relates to the mental ability to store and manage information required to complete real-time mental activities (maintenance and manipulation), e.g., reasoning, learning or understanding (Gathercole & Alloway, 2007).

One of the conventional method for assessing WM capacity is the digit span test, where participants are required to repeat digits from lists of increasing length, both in a forwards and a backwards order. The Digit Span Test, a subtest of the Wechsler-Adult Intelligence Test (Wechsler, 1955) was originally designed to assess participants' Intelligence Quotient (IQ). This technique has since been established in numerous studies as a reliable measure of working memory capacity (Gathercole & Alloway, 2007), and was therefore employed in the current study to control for participant differences with regard to this characteristic, which has been shown to be relevant to second language learning ability (Olsthoorn et al., 2014).

2.12 Conclusion and Summary

Findings from the literature review demonstrated the importance of vocabulary knowledge and reading in L1 as well as L2. Researchers have attempted to determine the threshold of vocabulary needed by a second language learner in order to be able to read adequately, but the percentage of vocabulary coverage required depends on how much comprehension of the text is necessary (Hsueh-chao & Nation, 2000; Laufer & Sim, 1985; Laufer, 1989; Schmitt et al., 2011). Through these studies, researchers have determined that a percentage of vocabulary coverage of more than 95% is needed to allow adequate comprehension of texts (Hsueh-chao & Nation, 2000). Moreover, although uptake of words

from incidental reading seems to be limited, reading can lead to vocabulary learning and enrich multiple aspects of word knowledge (Waring & Takaki, 2003; Horst, 2005; Pigada & Schmitt, 2006; Brown et al., 2008). These studies have also found that the frequency of word exposure contributes greatly to learning new vocabulary. Thus, vocabulary and reading in L2 seem to be correlated and share a bidirectional relationship: vocabulary coverage in the text affects reading comprehension performance, and, conversely, reading comprehension affects vocabulary learning and retention.

With reference to the importance of vocabulary knowledge as an important factor in L2 reading comprehension, researchers have provided empirical evidence for the contribution of depth of vocabulary knowledge to reading comprehension (Rashidi & Khosravi, 2010). Others emphasise, however, the importance of vocabulary size in order to achieve successful reading (Laufer, 1992; Hsueh-chao & Nation, 2000). The relative contributions of these two dimensions therefore need to be investigated further.

Various related models in L2 lexicon development were studied in the literature, of which one is important for the current thesis. Jiang (2000) proposed a three-stage model of lexical development in L2 learning which provides a clear insight into how L2 learners process and organise aspects of L2 word knowledge. However, linking this model to the analysis of learning new words from reading can provide further contribution to this knowledge. Another relevant theoretical model is the Involvement Load Hypothesis proposed by Laufer and Hulstijn (2001), which contributes to the explanation of the ways in which incidental vocabulary learning can occur. It is claimed that processing these words more elaborately will result in achieving a higher retention than by processing these words less elaborately, which means paying less attention to a word's aspects. Further tests of this hypothesis are needed to find out how we can increase students' engagement with words through elaboration activities.

The current research attempts to fill the gap in our knowledge about the relationship between vocabulary knowledge and reading comprehension performance. In particular, it focuses on the aspects of vocabulary knowledge that contribute to reading comprehension in a second language (Study 1), and, conversely, on the ways in which word learning and retention from reading in L2 can be enhanced (Study 2).

Chapter 3: Study 1

3.1 Aims of the Study

Study 1 investigates the relationship between vocabulary knowledge and reading comprehension. More specifically, the aim is, first of all, to establish to what extent reading contributes to vocabulary learning and retention in L1 Arabic learners of English, and, second, how depth and size of vocabulary knowledge contribute to reading comprehension.

The first aim of Study 1 is based on Jiang's (2000) three-stage model of lexical development in a second language. This model proposes that L2 learners who learn a second language in formal classroom settings do not acquire words in the same ways as L1 learners, because they do not have access to target language input that is sufficiently rich and because they are already in possession of fully-fledged L1 lexical and semantic systems. According to Jiang (2000), L2 learners generally start their L2 learning path by building a lexical entry associated with formal specifications only (phonetic/phonological properties). Thus, at this initial stage, L2 learners learn the spelling and/or pronunciation of the L2 word but not the L2 meaning or grammatical information associated with this entry. Instead, they create a link to the L1 lemma information through the translation equivalents. In the next stage, they transfer the L1 lemma (semantic and syntactic) information into the L2 lexical entry. In the first two stages, they tend to rely on the L1 semantic and syntactic system in learning new words. This means that they learn the form of the new words first associated with the L1 meaning. Only in the last stage do they integrate L2 semantic, syntactic and morphological information into the new L2 lexical entry. Jiang also points out that most of the L2 words are fossilised in the second stage (the L1 lemma mediation stage).

Jiang's (2000) model provides a new framework of reference within which we can analyse depth of vocabulary knowledge. Depth of word knowledge means how well L2 learners know different aspects of the form, meaning, and use of words. In the light of Jiang's model, it is to be expected that the form of a word in L2 is more easily recognised than the meaning, as the information from the L1 lemma blocks access to L2 semantic and syntactic information. This model, however, has not yet been applied to studies of the link between vocabulary depth and reading comprehension. In this sense, the current project tries to make an important contribution to knowledge by applying Jiang's (2000) model to the analysis of vocabulary learning from reading.

Read (2004) points out that the construct of depth of vocabulary knowledge is less well studied than vocabulary size, because investigating depth of word knowledge is more

complicated than the vocabulary size construct. Investigating vocabulary depth further is important because it is not well known to what extent vocabulary knowledge contributes to reading comprehension. Having reviewed the related literature on reading comprehension, researchers have provided empirical evidence for the importance of depth of vocabulary knowledge in reading comprehension (Rashidi & Khosravi, 2010), however, there is much less on the relationship between vocabulary and reading. Although many researchers agree that size and depth are shown to be closely correlated (Qian, 1999; 2002; Li & Kirby, 2014), the relative contribution of each to reading comprehension remains unclear, in part due to differences in the ways in which researchers operationalise these dimensions.

Word frequency plays a central role in vocabulary learning because it determines the words that learners are likely to encounter and how often those words are encountered. The relationship between word frequency and the likelihood of their occurrence and the ability to learn them is found to be strong (Milton, 2007). Researchers (Horst et al., 1998; Pigada & Schmitt, 2006; Brown et al., 2008) found that the frequency of word exposure is important to the learning of new words and leads to developments in multiple aspects of word knowledge. Waring and Takaki (2003) showed, however, that even if new words were repeated 15 to 18 times, there was less than a 10% chance that these were successfully translated three months later, and no words repeated fewer than five times were translated correctly. This clearly shows that although higher repetition rates are valuable, learning a word may depend on more than only frequency of exposure.

Study 1, then, aims to explore which aspects of word knowledge (form, meaning, and use of words) L2 learners acquire and retain through reading comprehension. An additional objective for this study is to identify the extent to which the learner's vocabulary size or vocabulary depth can explain variance in reading comprehension performance. A third main aim for this study is to find out whether the frequency of occurrence of target non-words in a text (i.e., target words appeared either four or eight times in the passages), or the frequency band (high or low) of the original words which were replaced by the target non-words can explain learning and retention of these words. In this study an interview was conducted to evaluate participants' familiarity with the target non-words, and L1 English native speakers were also compared with L2 learners of English in terms of their word learning and retention performance.

3.2 Research Questions, Hypotheses and Predictions

Study 1 sought to answer the following two research questions:

RQ1a: Which aspects of word knowledge (form, meaning, and use) of new words encountered in a reading comprehension task will L2 readers retain?

Hypothesis: The form of new L2 words will be retained better than their meaning or aspects of use (in a sentence) (Jiang, 2000).

Prediction 1: Better comprehension will be associated with better retention of word meaning and use (in a sentence).

Prediction 2: Better comprehension will be associated with less decay in form recognition.

RQ1b: How do participants determine the meaning of the four target non-words while reading the texts? (qualitative feedback from participants)

RQ2a: To what extent can vocabulary size and vocabulary depth explain variance in reading comprehension?

Hypothesis: Vocabulary depth will explain additional variance in reading comprehension over and above the variance explained by vocabulary size (Rashidi & Khosravi, 2010).

Prediction 1: The association between vocabulary depth and reading comprehension questions will be stronger than that between vocabulary size and reading comprehension questions.

Prediction 2: The association between vocabulary depth and reading summarisation questions will be stronger than that between vocabulary size and reading summarisation questions.

RQ2b: To what extent are the new words selected for inclusion in the texts appropriate in that participants for whom English is the L1 can derive the meaning of these new words from the texts?

Prediction 1: The patterns of performance measures will be generally similar across the L1 and L2 groups, in terms of their word learning and retention from reading, but that any differences will reveal anomalies in the task materials.

RQ3: To what extent can frequency of occurrence and frequency bands of target words explain learning and retention of these words encountered through reading?

Hypothesis: The frequency of occurrence of a number of new words in a text will explain learning and retention of these words (Pigada & Schmitt, 2006).

Prediction 1: New words which occur more frequently in the text will be better learned and retained.

Prediction 2: New words which replace high-frequency words in the text will be better learned and retained.

3.3 Methods

3.3.1 Participants

A total of 60 participants took part in this study: 30 L1 Arabic learners of English from Qassim University in Saudi Arabia and 30 native English speakers from Reading University. The University of Reading participants were all native English speakers (27 females and 3 males) and enrolled in undergraduate classes. Their ages ranged between 21 and 30 years, and the mean age was 23.9.

All L1 Arabic learners of English were females, their ages ranged between 18 and 20, and the mean age was 19.2. All of them had been exposed to English for a period ranging between 9 and 11 years (see Appendix 1). English exposure for these L1 Arabic learners of English was in a formal instructional setting. All participants were enrolled in the Qassim University Preparatory Year Program, available on *Qassim University* website, in the English Language Department. The programme includes four components: Maths, Computing, Arabic and English, and lasts for one full year. It is divided into two levels: A1-A2 (in the first academic semester) and B1-B2 (in the second academic semester). The division of the preparatory year into Levels A and B parallels the Common European Framework of Reference for Languages (CEFR) levels, as the program offers Unlock, a five-level academic skills course accredited by Cambridge University Press, which is based on the CEFR to indicate the levels of its resources. Upon completion of Level B, participants are expected to enrol in their respective program majors, provided they pass Level B examinations in English as well as in all other components. The passing grade is 60%. Data collection for this study took place during the second academic semester of the 2015 academic year, which means that participants' English proficiency levels were expected to be between pre-intermediate (B1) and upper intermediate (B2), in accordance with CEFR levels.

The present study has not attempted to take into consideration L1 reading ability as a factor which may affect reading comprehension assessment for two reasons. Firstly, it can be argued, as earlier research found, that L2 reading ability is likely to be affected by L2 language ability more than L1 reading ability (Cziko, 1980; Devine, 1987; Jiang, 2011; Jeon & Yamashita, 2014). According to Clarke (1979), those who can read well in their L1 may fail to transfer

their reading ability to their reading in L2 due to limitations in their L2 proficiency. Carrell (1991) further proposes that L2 language ability is more important than L1 reading ability in L2 reading, especially for L2 learners at lower proficiency levels. Secondly, no L2 reading test was included for reasons of practicality, as there is no Arabic reading proficiency test that is accepted as a standard, and there is no recognised equivalent for the TOEFL or IELTS in Arabic. Different Ministries of Education across the Arab world use their own "in-country" examinations, and the main matriculation examination takes place at the end of secondary school (Secondary School Leaving Certificate). Students are expected to be able to read with understanding texts of a general, academic and literary nature. L2 learners participating in this study from Qassim University were at this level.

3.3.2 Research instruments

In order to examine the participants' vocabulary size and depth of vocabulary knowledge, two offline tests, the Vocabulary Size Test (VST) (Nation & Beglar, 2007) (see Appendixes 2a and 2b) and the Vocabulary Knowledge Scale (VKS) (Brown, 2008) (see Appendix 3), modified from Wesche and Paribakht (1996), were used. A Lexical Decision Task (LDT) was designed by the researcher to assess online the accuracy and speed of participants' recognition of the form of target words, and was conducted in two test sessions: immediate post-test (IPT) and delayed post-test (DPT), administered after two weeks. The York Assessment of Reading for Comprehension (YARC) Secondary Test, published by Stothard et al. (2010), was used to test the participants' reading comprehension performance which includes reading comprehension questions (RCQ) and reading summarisation questions (RSQ) (see Appendixes 4a – 4l). To evaluate participants' familiarity with the target non-words, an interview was conducted after they had taken part in the study (see Appendix 5). Each of these instruments will be presented in more detail from Section 3.3.4 to Section 3.3.8.

3.3.3 Overall design

The current study was developed as a correlational study, which investigates in detail the relationship between vocabulary knowledge and reading comprehension performance. To provide a baseline, the performance of L1 English speakers was measured to ensure that there were no aspects of the passages or non-word targets that would present problems for L1 speakers. Any consistent instances of sub-optimal performance in the L1 sample would mean the test would need to be revised. Two groups, an English native speaker group (G1) and Arabic learners of English (G2), participated in the present study. The study, further, used a cross-sectional design in which two groups were compared at one point in time. A detailed description

of the dependent and independent variables and how they map on to research questions will be given in the following sub-sections.

3.3.3.1 Variables relating to RQ1a

The independent variables were the G2 YARC reading comprehension (RCQ) scores and reading summarisation (RSQ) scores. The VKS target non-word scores and the target non-word accuracy scores of the LDT (IPT and DPT) of the G2 group were the dependent variable (see Table 3.1).

3.3.3.2 Variables relating to RQ2a and b

The principal independent variables for RQ2a were the accuracy scores from the VST, VKS. Scores on the YARC reading comprehension (RCQ) and reading summarisation (RSQ) for both passages (A and B) were the dependent variables. For the RQ2b, the group variable was the independent variable, and the scores from the VKS, and the LDT (IPT and DPT, including accuracy scores and Reaction Time) were the dependent variables (see Table 3.1).

3.3.3.3 Variables relating to RQ3

Frequency of occurrence of non-words in the texts (four versus eight occurrences) and the frequency of original items which were replaced by the non-words in the texts (high versus low) were the independent variables. Scores on the VKS, and LDT (IPT and DPT, including accuracy scores and Reaction Time), were the dependent variables.

Table 3.1 The overall design of Study 1

RQ	Prediction	Group	Independent Variables	Dependent Variables
1a	1	Only G2 scores	RCQ and RSQ for both passages A and B	The target non-word accuracy scores of LDT (IPT and DPT)
1a	2	Only G2 scores	RCQ and RSQ for both passages A and B	The target non-word scores of VKS
1b	Qualitative analysis of participants' responses			
2a	1	G1 and G2 scores were combined	The VKS and VST scores	RCQ for both passages A and B
2a	2	G1 and G2 scores were combined	The VKS and VST scores	RSQ for both passages A and B
2b		G1 and G2 were compared	The group variable	VKS scores and LDT accuracy scores and RT (IPT and DPT)
3	1	G1 and G2 were compared	Frequency of occurrence of non-words in the text (four versus eight occurrences)	VKS target non-word scores and LDT accuracy target non-word scores and RT (IPT and DPT)
3	2	G1 and G2 were compared	Frequency of original items in the text (high versus low)	VKS target non-word scores and LDT accuracy target non-word scores and RT (IPT and DPT)

3.3.4 Vocabulary Size Test

The VST developed by Nation and Beglar (2007) was used in the current research as a receptive vocabulary size measure. It was developed to provide a reliable, precise, and comprehensive instrument of L2 English learners' written receptive vocabulary size (Beglar, 2010). It consists of fourteen 1,000 word family frequency lists derived from the BNC (Nation, 2006), and the test samples 10 items from each frequency level, totalling 140 items with 14

levels of difficulty, becoming incrementally higher in a meaningful order. It has a multiple-choice test format that presents each target word within a short non-defining sentence. The VST can be an appropriate research tool for studies based on learners who are studying in an English medium system and have English as their second language (Nguyen & Nation, 2011). It has been successfully conducted for research purposes and found to be a valid and reliable test (e.g., Beglar, 2010; Nguyen & Nation, 2011).

By conducting a Rasch-based validation of the VST, Beglar (2010) found that it has some good technical features, which makes it an appropriate measure for use in the current study. These features are suitable for learners of different proficiency levels, because they measure what is expected to be measured (the written receptive vocabulary knowledge required for reading); they have a wide range of clear and unambiguous items which become progressively more challenging as the frequency levels decrease, in order to clearly distinguish different levels of vocabulary knowledge; they can be administered in two halves so that learners do not need to sit all 140 items of the test but a 70-item version only; and it is easy to score and interpret. Beglar (2010) further showed that even when comparing learners with different proficiency levels, comparing male participants with females, and comparing the 70-item version of the test with the 140-item version, the VST performs consistently and reliably (reported as > 0.96). It seems, therefore, that Beglar's (2010) study provides considerable evidence for the validity of the VST.

An additional reason to justify the choice of the VST, as mentioned in the literature review (Section 2.11.3.2), is that it includes a range of frequency levels representing sample words within their frequency bands. In comparing the VST with the VLT (see Chapter 2, Sections 2.11.3.1 and 2.11.3.2), the VLT test levels consist of words from the 2,000, 3,000, 5,000 and 10,000-word levels, which means that the tenth word level estimates learners' vocabulary size between 5,000 and 10,000 words. In Laufer (2000), the estimated vocabulary size of educated L2 learners of English ranged from 1,000 - 4,000 word families. However, there is a considerable variability in vocabulary knowledge among L2 learners. Alsaif (2011), for example, found that the English vocabulary size of his L1 Arabic adult students was below 1,000 words. Thus, the VST might be the appropriate choice to be administered in the present study, since it is broken into 1,000 frequency bands, which makes it useful in the current study, particularly for L2 participants of B1 and B2 English proficiency levels to measure their vocabulary size in the first five frequency bands (1k-5k). All the items used in this test have been carefully devised and trialed (Nation, 2011), and hence, it was employed in the current study to measure participants' written receptive vocabulary size.

3.3.4.1 Administering and scoring the test

In this study, not all the levels were used to estimate students' vocabulary size. In consideration of L2 learners' vocabulary level, only the words from the first 1,000 to the eighth 1,000 were used out of the total word range to the fourteenth 1,000 (see Appendixes 2a and 2b). The other levels were omitted because, as stated by (Laufer, 2000), a reading target of 3,000 to 5,000 word families can be attainable for L2 learners, whereas the 8,000 to 9,000 targets can be difficult to achieve. Hence, including too many unknown words would have been demotivating for the participants in the current study, particularly for L2 participants of B1 and B2 English proficiency levels. The presentation order of the frequency levels of the test was randomised, in addition, the order was counterbalanced across participants (half of the participants took the test with order A and the other half took the same test with order B) to minimise order effects (see Appendixes 2a and 2b). Following Nation and Gu's (2007) VST scoring procedure, the participants' scores of the 10 items from each 1,000-word level was multiplied by 100 to calculate their total receptive vocabulary size. For instance, if a learner in the present study scored 35 out of all 80 items, the total vocabulary size was 3,500 words.

3.3.5 The York Assessment of Reading for Comprehension

The York Assessment of Reading for Comprehension (YARC), published by Stothard et al. (2010), consists of two testing levels: Primary and Secondary. The YARC Secondary Test is designed for assessing the reading comprehension of native speakers of English aged 11 to 16 years. It is based on the SVR model (Gough & Tunmer, 1986), in which reading comprehension consists of two skills: decoding and linguistic comprehension (see Chapter 2, Section 2.2). This means that learners need to master these two skills in order to be good readers: the ability to recognise words in texts and then to understand the underlying concept. In the current study, the YARC consists of two sections: decoding and reading comprehension. The YARC reading passages are provided at different levels, ranging from below average to average, above average and excellent. Two sets of YARC passages are paired for each level: fiction (A) and non-fiction (B).

3.3.5.1 Reliability and validity

As shown in the test manual (Stothard et al., 2010), the result of reliability scores for the different levels of reading comprehension lies between 0.85 and 0.90, i.e., it ranges from good to excellent. A lower reliability assessment, however, has been reported for summarisation skills, lying between 0.65 and 0.74. For this reason, the test developers recommended that an estimate of participants' reading comprehension skill should be based on two passages of

appropriate difficulty level, a practice that was employed in the current study. Furthermore, the second passage was needed in the present study to ensure there would be an adequate number of target items for readers to learn from reading (two in each passage), as both passages were used to measure learning and retention of non-words and the targets were different for each passage.

In considering the YARC's content validity, the test developers administered the validity of reading comprehension questions from both forms fiction (A) and non-fiction (B) to investigate whether these questions are a valid measure of reading comprehension skills or they can be answered from general knowledge. Students in Years 8, 9 and 10 from three secondary schools were asked to answer the comprehension questions from both forms A and B of YARC without reading the passages. They were instructed that these questions were about a reading passage that they had not read and they need to guess the answers. In general, these students were unable to respond correctly to the reading comprehension questions (Stothard et al., 2010), hence, YARC comprehension questions of forms A and B are considered as having a valid measure of reading comprehension skills. Concurrent validity is assessed by computing the correlations between the summarisation score and comprehension score for each level. The results show a moderate correlation of between 0.58 and 0.74 for the summaries and responses to the comprehension questions. The YARC is, therefore, considered to have acceptable concurrent validity. Thus, the YARC Secondary Test was adopted for this project in order to assess participants' decoding, which measures reading accuracy through the Single Word Reading Test (SWRT), and reading comprehension. These two components will be addressed in the following sections.

3.3.5.2 Single Word Reading Test

The Single Word Reading Test (SWRT) was used to select the appropriate level of difficulty of the YARC Passage Reading test for the participants. It assesses learners' ability to decode words and their sight reading skills. It consists of 70 words in total (see Appendix 4a). One point was awarded for each word read correctly and zero points for each word read incorrectly. No response (NR) was given if the student failed to answer.

3.3.5.3 Suitability of YARC for L1 Arabic learners of English

This YARC Secondary Test was originally designed for children English native speakers aged 11 to 16 years old, however, it can be used with students who learn English as an additional language (EAL). Stothard et al. (2010, p. 114) report that EAL students participated in a test to

ensure that the sample was representative of UK school students. It is worth noting that in the current study, this test is used for the first time by L1 Arabic learners of English. As it is well known that tests that were developed for monolinguals are not necessarily suitable for bilinguals or second language learners (Gathercole, 2013a; 2013b), it is important to carefully monitor whether the test is suitable for L1 Arabic adult students of English. Stothard et al. (2010) report that the SWRT scores provide a reading age score for children whose first language is English (see Appendix 4b). For identifying the suitability of the YARC Secondary Test for adult learners of English, a study conducted by Huang (2013) found that the age equivalent of Chinese adult students to English native speakers, in terms of their SWRT scores, was 9 years old. For reading comprehension, the age equivalent was found to be slightly higher, at 10 years and 7 months. Huang (2013), therefore, concluded that because the YARC Secondary Test was designed for native children aged between 11 to 16, Chinese adult students seem to lack the relevant abilities of SWRT and passage reading. Thus, Huang (2013) suggested that supplementary YARC passage, designed for students who have difficulties in reading and are unable to access levels of the YARC Secondary passages, would probably more suitable for Chinese adult students. Based on L2 participants' SWRT scores, L1 Arabic adult students study have an average reading age which is equivalent to a native speaker age of 11 years and three months (see Table 3.2 and Appendix 4b). Considering this, SWRT scores for L1 Arabic adult learners fell within the range for native English children, the YARC Secondary Test was considered suitable to be completed by L1 Arabic learners of English who participated in the current study.

Table 3.2 Mean accuracy scores and standard deviations of L1 Arabic students (G2) on the SWRT and age score, based on norms from children whose first language is English

SWRT (scores)		**SWRT**
Mean	**SD**	**(year)**
47.21 (67.44%)	2.97	11.03

3.3.5.4 Passage reading (A) and (B) including the target words (non-words)

In the present project, Level 1 of the YARC Passage Reading test was selected, based on L2 learners' raw scores (ranging from 43-51) on the SWRT. The same two passages were given for all participants (G1 and G2) to make sure that everyone was exposed to the same target

words. As part of the process of selecting the comprehension passages, the level of difficulty of the reading passages was also investigated using *the vocabulary profiler* (BNC-20 version), available on the *Compleat Lexical Tutor* website, which provides information about the frequency layers of the words in the text. The low-frequency words in the texts were replaced with higher frequency words (see Table 3.3) to ensure that the readers knew around 97.50% of the words in the text (Laufer & Sim, 1985; Nation, 2001) (see Appendix 4c and 4h). Hence, all words in the two texts were below 4k, since Laufer (2000) reported that the vocabulary size of educated second language learners of English often ranges from 1,000-4,000 word families. This way, it can be ascertained that reading the text would be challenging but not too complicated.

Target non-words in texts (A) and (B) were changed to non-words in order to replicate the design reported by Waring and Takaki (2003) and Brown et al. (2008). The decision to use non-words rather than real words was made to ensure that they were totally unknown prior to the reading and to ensure participants would not be exposed to these words in between the immediate and delayed post-test, so that all learners needed to acquire the form and meaning of the words from the reading passage and no one had an advantage over the others because of pre-existing knowledge. It would also ensure that their scores on the post-test after the delay would not be affected by their regular language studies.

It was necessary to decide what types of words (e.g., nouns, verbs or adjectives) should be chosen and how many words would be targeted. Nouns in the singular were selected, as they are easier to identify than adjectives and adverbs (Laufer, 1997b). Verbs were not chosen, as they appear with person and tense inflections and therefore could be harder to identify (Brown et al., 2008). A glossary of definitions was not given during the test.

The target non-words were created carefully according to Arabic rules in order to suit Arabic learners of English. In order to eliminate any potential matching of these non-words to either Arabic or English words, they were tested by educated English speakers and Arabic speakers in a pilot study (see Appendixes 6a and 6b). As discussed in Chapter 2 (Section 2.10.1), Arabic words are based on a root that contains three consonants, and these three consonants can be combined with other vowel patterns to produce a family of lexical items sharing a common meaning. As an example, the root *d-r-s* joins with different patterns of vowels to create *madrasa* "school," e.g., *darrasa* "to learn" (Ryan & Meara, 1991). The target non-words were therefore generated with the following pattern: consonant/vowel/consonant/vowel/consonant/vowel, an example being the Arabic word *kataba*

"to write," in order to help Arabic learners of English to process and become familiar with the non-words.

The second passage was needed to ensure there would be a sufficient number of target items for readers to learn from reading, as the two passages were used to measure learning and retention of non-words per story. It was not possible to include more non-words in one passage because the density of unknown words in the text would be too high. As it was necessary to use a text in which 97.5% of the words were already known, 2.5% of the text (approximately 12 target words for each passage) was composed of target non-words. Nation (2001) notes that if unknown words appear within a familiar context, this might facilitate acquisition. This level of coverage should be ideal for determining the participants' ability to identify the meaning of unknown words from their context (Laufer & Sim, 1985; Hsueh-chao & Nation, 2000).

The frequency of occurrence of the words was controlled because this is known to affect a learner's chances of acquiring a new word (Pellicer-Sanchez & Schmitt, 2010; Eckerth & Tavakoli, 2012). It was unclear how many occurrences were needed, and estimates vary from study to study. The total number of target non-words in each text, then, was twelve from the original text of approximately 480 words, which were categorised into two sets of identified words. In text (A): *cenedies,* a target non-word, replaced a high-frequency (1k) existing real word "human" which occurred eight times in the text. Another target non-word, *toroko,* replaced a low-frequency real word (4k) "sting" which occurred four times in the text and is more difficult to identify. In text (B), there were also two types of target non-words: *sataca,* a non-word, replaced a high-frequency word (1k) "trip" which appeared in the text four times and was harder to identify, and another non-word, *pocoko,* replaced a low-frequency real word (3k) "suburb" which occurred eight times in the text and was easy to identify (see Table 3.3).

Table 3.3 The categorisation of the target non-words in the texts (A & B)

Passage A			**Passage B**		
Target non-words	**Frequency level**	**No. of appearances in the text**	**Target non-words**	**Frequency level**	**No. of appearances in the text**
cenedies	1k	8 times	*sataca*	1k	4 times
toroko	4k	4 times	*pocoko*	3k	8 times

Table 3.4 A list of the low-frequency words in the texts that were replaced with higher-frequency words

Passage A		Passage B	
Original words	**Replaced words**	**Original words**	**Replaced words**
beak	nose	suburb (target word)	pocoko
drab	colourless	trip (target word)	sataca
fake	light	harsh	hard
beehives	bees' nests	submerge	fill
sting (target word)	toroko	adjacent	next
fatal	kill	deliberately	intentionally
humans (target word)	cenedies	invaded	entered
elegant	beautiful	villagers	people
territory	county	peal	ring
grass	green	resonated	vibrated
helpful humans	helpers	consistency	flexibility
chattering	non-stop	fade	disappear
equipped	fitted	_____	_____
hive	nest	_____	_____
honeycombs	honey pieces	_____	_____
please	want	_____	_____
tribes	families	_____	_____
greedy	hungry	_____	_____
poisonous	harmful	_____	_____
ungrateful	thankless	_____	_____

3.3.5.5 Reading comprehension questions and summarisation question

In the present study, the reading comprehension questions were selected from thirteen YARC comprehension questions for each text, on the basis of which were most relevant to the key ideas of the passage, and, furthermore, most relevant to the target non-words. There were four comprehension questions for each of Passage (A) and Passage (B). Two of the reading comprehension questions for each passage targeted the non-words in the text (see Appendixes 4d and 4i). Each correct answer to the reading comprehension questions was worth one point (see Appendixes 4e and 4j), and then the total number of the correct answers was computed.

In addition to the four comprehension questions, participants were asked to perform a final summarisation question for each passage, as specified in the YARC test manual (Stothard et al., 2010). The purpose of the summarisation question was to provide a further measure of reading comprehension in addition to the reading comprehension questions. The summarisation question required participants to summarise the content of the passage they had just read. Its function was to assess the participants' capacity to identify the key points of the passage (see Appendixes 4f and 4k). Following the scoring guidelines mentioned in the test manual (Stothard et al., 2010), each correct item provided in the summary was worth one point, and then the total number of summarisation items correctly recalled was computed (see Appendixes 4g and 4l). The summarised items do not need to be recalled in the same order as presented in the passage. Following the instruction mentioned in the YARC manual (Stothard et al., 2010), participants' answers to the comprehension questions and the summarisation questions were recorded using an audio recorder. Their answers were then transcribed by the researcher verbatim for the purposes of scoring.

3.3.5.6 Scoring procedures

In the scoring of the YARC Passage Reading assessment, dialect and accent differences were not marked as errors. Moreover, self-corrections (when participants spontaneously corrected themselves) were not counted as errors. However, reversals of words (e.g., *was* reversed to *saw*) were marked as errors. Additionally, inserting a word or part of a word (e.g., it was *raining* – it was *not raining*) was counted as an error. When participants were unable to attempt the word in the decoding test, the researcher waited for approximately five seconds before supplying the correct word.

3.3.6 Interview with participants

To evaluate participants' familiarity with the target non-words, an interview was conducted after they had taken part in the study, two weeks later. The participants were asked how they determined the meaning of the four target non-words while reading the texts (see Appendix 5). Participants' answers were recorded using an audio recorder. Their answers were then transcribed by the researcher.

3.3.7 Lexical Decision Task

The Lexical Decision Task (LDT) is considered to be one of the most widely used paradigms for measuring word recognition processes and investigating the structure of the mental lexicon (Elgort, 2011). Because the first research question of this study focused on

understanding the processing of L2 lexical items (Jiang, 2000), it was important in this study to apply an online test which could measure lexical access and visual word recognition processes. As mentioned in Chapter 2 (Section 2.11.5), these online measures allow the researcher to obtain information about accuracy of word identification, error rates and reaction times (RTs) (Elgort, 2011). Li and Kirby (2014) point out that a single test cannot completely assess all aspects of word knowledge, which means a combination of measures is needed for more precise assessment. In the present study, an LDT was developed to measure recognition of word form (accuracy and speed).

3.3.7.1 Administering and scoring the test

In LDT tasks, respondents are normally asked whether or not the target words are existing English words, but this question was not appropriate in the current study, where respondents had learned non-words. Therefore, participants in this study were required to decide whether they had seen the word before as quickly and as accurately as possible. Reaction times were measured as an indication of the speed of processing the non-words. The items in the LDT were categorised into four groups: target non-words, non-words not occurring in the two texts, real words from the texts, and real words not occurring in the two texts (see Table 3.5). Different categorised items were used because in this type of experimental paradigm, the number of distractors has to be large enough so that participants' attention is not drawn to the target items too much. Real words may also be too easily identifiable, so real words with the cvcvcv structure of the target non-words, such as *potato*, *computer*, and *manager*, were used as additional distractors.

For each passage, the LDT consisted of 32 items: 16 non-words that did not occur in the text and 16 real words providing a balance of distractors. The 16 real-word category contained the following: 2 target non-words; 6 six- or seven-letter real words from the text (three high frequency and three low frequency); 4 six- or seven-letter real words that did not occur in the text (two high frequency and two low frequency); and 4 six-letter real words that did not occur in the text and had a cvcvcv structure (two high frequency and two low frequency) (see Appendix 7).

Table 3.5 The items in the LDT (IPT) for each passage

	Real words	**Non-words**
From the text	6	2 (target words)
Not occurring in the text	4	16
Not occurring in the text with a cvcvcv structure	4	_____

The LDT in the post-test (DPT) contained 64 items: 32 non-words that did not occur in the text and 32 real words (including the four target non-words) which were treated as "real words" for the purpose of the LDT. The remaining 28 real words were 12 six- or seven-letter real words from the texts (six high-frequency and six low-frequency); 8 six- or seven-letter real words that did not occur in the texts (four high frequency and four low frequency); and 8 six-letter real words that did not occur in the texts, with a cvcvcv structure (four high frequency and four low frequency) (see Table 3.6). Because the LDT was administered three times, only the target words were repeated each time, but the distractors (non-words not from the texts, the real words not from the texts with or without a cvcvcv structure) were changed in each administration. This procedure was followed in order to avoid repetition of test items across tests, which may have increased the chances of potential learning effects for words that were not target items in the study.

Table 3.6 The items in the LDT (DPT) from both passages (A & B)

	Real words	**Non-words**
From the texts	12	4 (target words)
Not occurring in the texts	8	32
Not occurring in the texts, with a cvcvcv structure	8	_____

All the LDT stimuli were nearly of the same length and presented individually in the middle of the screen using black 18-point Times New Roman font against a white background. Two presentation orders had been prepared for each experiment; half the participants were presented with all items in one order and the other half were presented with items in the reverse order. The two list presentations were used equally across the participants in the sequence in order to avoid any inter-stimulus effects between the sequence of words in the list that may affect participants' performance. The items in the lists were also systematically pseudo-randomised so that the target non-words occurred at least ten items apart and there were not

more than three non-words adjacent to one another and not more than three real words following each other.

The participants were instructed as follows: “Words will appear individually in the middle of your screen. You should decide whether you have seen each word before as quickly and as accurately as possible. Press 'M' if yes (seen before) and press 'X' if no (not seen before). You will see a '*' in the centre of the screen for 20 seconds before the first word is presented. During this time, please place a finger of your left hand on the 'X' key and a finger of your right hand on the 'M' key, so that you are ready to start when the '*' disappears. You will have 6 practice trials to familiarise you with the procedure. Press the 'Space Bar’ when you are ready to start” (see Appendix 8). After that, at the beginning of the experiment, six stimuli different from those selected for the main experiment (three real words and three non-words) were presented as practice trials to make sure that participants became familiarised with the task and that their reaction times were valid from the first experimental trial. Any participant whose error rate was higher than 40% in the practice trial was rejected and replaced by a new participant. The participants were awarded one score for each item answered correctly. All error trials were excluded. In order to ensure that extreme reaction time data did not affect the mean RTs for the LDT, an outlier identification procedure was followed whereby RTs that were more than two standard deviations (SD) above or below the mean for any participant were trimmed to the cut-off value of two SDs for that participant. A similar outlier procedure is followed in Forster and Veres (1998), McRae and Boisvert (1998), and Elgort (2011).

3.3.8 Vocabulary Knowledge Scale

Depth of word knowledge was measured using Wesche and Paribakht’s (1996) Vocabulary Knowledge Scale (VKS). This test is frequently used by researchers interested in the “incidental” learning of new words through reading activities (Read, 2007; Jahangiri & Abilipour, 2014), Others have used modified versions of the VKS in L2 vocabulary development research (Joe, 1998; Zareva, Schwanenflugel & Nikolova, 2005; Brown, 2008). As mentioned in Chapter 2 (Section 2.11.4.2), the original VKS was compiled by Wesche and Paribakht (1996), who devised a five-point elicitation scale from (1) “I don’t remember having seen this word before” to (5) “I can use this word in a sentence.” The VKS scale combines self-report with some demonstrable evidence of word knowledge, shown as a synonym, L1 translation or sentence (Read, 2007). As pointed out by Tan, Pandian and Jaganathan (2016), there are provisions in the VKS test which prompt the test takers to show their receptive knowledge of a word as well as their capacity to use it productively. Therefore, vocabulary

tests, such as the WAT and the DVK, which focus on form-meaning without testing for production, remain essentially assessments that measure receptive word knowledge only.

In terms of testing the VKS for reliability, in their research involving a university-level ESL of different proficiency levels, Wesche and Paribakht (1996) found high correlations (above 0.90) between learners' self-report of word knowledge and the actual score for demonstrated knowledge of the word. The test-retest procedure revealed reliability measures of above 0.80. Paribakht and Wesche (1997) describe the VKS as a practical measurement that can be used with any words and is useful for research focused on recognition and use of words. The VKS was appropriate for the design of the present study to measure the depth of word knowledge (Tan et al., 2016), since the objective was to examine lexical ability in a more productive and comprehensive way or to have a clear picture of the effects of a vocabulary learning intervention. A two-dimensional vocabulary test, such as the VKS (Wesche & Paribakht, 1996) is therefore an appropriate choice. Another reason justifying the choice of the VKS is that it assesses the three aspects of word knowledge that are the focus of the current research, namely, form, meaning, and use in a sentence. Tests such as the WAT and the DVK, as mentioned in Chapter 2 (Section 2.11.4.1), assess only the meaning and collocation of the target word, regardless of other aspects of word knowledge, such as using the target word in a sentence. Furthermore, researchers (e.g., Bogaards, 2000; Greidanus, Beks, & Wakely, 2005) who developed the WAT, found this test to be challenging for even advanced EFL learners at university level. It seems, therefore, that the WAT might be more appropriate for advanced L2 learners but not for lower level learners, as in the case of the L2 participants in the current study whose English proficiency levels were between B1 and B2. The VKS, by contrast, is not complex, and was therefore deemed to be more appropriate for L2 participants in the current study.

Over the years, the VKS has been used frequently in ESL vocabulary knowledge testing, and the scale test and its modified versions have been conducted in various studies (Waring, 2002). In order to measure word knowledge more accurately, the original 5-point VKS scale (Wesche & Paribakht, 1996) was simplified by Brown (2008) to a 4-point scale. As Bruton (2009) points out, the original VKS scale developed by Wesche and Paribakht (1996) involved two statements which may not be clearly differentiated: "I think it means" (Scale III) and "I know this word. It means" (Scale IV). Therefore, the following statement: "I have seen this word before and I think it means _____' (Scale III) was removed in the modified version of the VKS developed by Brown (2008). This simplified updated version of the VKS (Brown,

2008) was used in the present study (see Table 3.7), as it consists of a -point scale clearly divided into two independent tests of receptive and productive word knowledge**.**

Table 3.7 Modified Vocabulary Knowledge Scale

	Self-Report Category
I.	I don't remember having seen this word before.
II.	I've seen this word/phrase before, but I don't know what it means.
III.	I know what this word. It means _____ (Give the meaning in English or Arabic.)
IV.	I can use this word in a sentence _____ (Write a sentence.) (If you do this section, please also complete III.)

3.3.8.1 Administering and scoring the test

The VKS (Brown, 2008) comprises a four-point scale to capture specific stages in the initial development of core knowledge (see Table 3.7). The unknown word category includes unknown words (Category I) and partial knowledge (form) (Category II). The known word category includes receptive word knowledge (meaning) (Category III) and productive word knowledge (use) (Category IV). The scale ranges from I, which represents total unfamiliarity, to IV, which represents the capacity to use a word grammatically and semantically accurately in a sentence. In the present study, the participants took the VKS for all four target non-words from (A) and (B), combined with six real words to assess the meaning and use aspects of those words retained over time (see Appendix 3).

There were no points given to the unknown word category, Category I: "I don't remember having seen this word before." Furthermore, as in Baleghizadeh and Abbasi's (2013) study, no points were given to Category II: "I have seen this word before but I don't know what it means." As in the case of the current investigation, only the written receptive and productive word knowledge was tapped by the VKS, since the LDT, used before the VKS, had already measured form recognition. That is, if learners chose Category I or II, they did not gain any points. Participants were awarded two points in the known word category: one point for receptive word knowledge if a synonym or translation of the target word was accurate. A wrong synonym or translation response was, however, given zero points. Learners were also awarded one point for productive word knowledge when the use of a target word was both grammatically and semantically accurate.

3.3.9 Pilot study

Before administering the main experiment, a small pilot was conducted with five L1 Arabic learners of English and five English native speakers who were selected from the University of Reading campus community. The same methods and procedures were employed for both groups of participants. The YARC test, the VST, LDT and VKS were piloted to ensure that the experimental procedure was appropriate for English native speakers and L1 Arabic learners of English. Following the pilot, both groups of participants were asked to provide feedback about the experiment to determine if there were any aspects of the test design that needed to be improved. No changes were made as a result of their feedback. This was essential to ensure that the design of the study was valid.

3.3.10 Procedure

The participants were not told what the main purpose of the study was: to avoid intentional new vocabulary learning. They were told that the study focused on reading comprehension in Arabic learners of English and that groups of L1 English speakers and L1 Arabic learners of English were needed for the study (see Appendix 9). The participants firstly completed a brief questionnaire with information about their educational background, age and interests, since these factors may affect L2 learners' performance (see Appendix 1). They then received the decoding test (SWRT) to establish whether the YARC passage selected was at an appropriate level for all participants. The researcher instructed the participants to read each word by themselves, and, where the learner did not know the word, s/he was encouraged to try to sound out the word; if the learner still did not know the word, they then moved on to the next word. Learners read the words as accurately as possible. The researcher was also allowed to help participants to read the words or tell them whether each word was right or wrong. The participants were told that their responses would not be returned to them, but they would receive a summary of the overall results. They were then given the VST (Nation & Beglar, 2007) in order to test their vocabulary size.

A day later, the YARC Passage Reading (A), containing two non-words, was administered. The participants were instructed to treat any novel words in the text as real words. Immediately after reading passage A, the participants took the YARC comprehension test on that passage, which comprised four comprehension questions and a final summarisation question. Participants were allowed to refer back to the passage to answer the four comprehension questions. However, they were not allowed to return to the passage to answer the final summarisation question. Then, they took the LDT to measure the form recognition of the words.

A day later, the subjects read the YARC Passage Reading (B) containing two non-words. The same procedures for passage A were repeated for passage B (see Figure 3.1).

Two weeks later, the participants took the VKS of all non-words from passage (A) and passage (B) to assess retention of the meaning and use of the words, and the LDT to assess word recognition performance after a two-week delay. A quick interview was also conducted with the participants to assess their familiarity with the target non-words (see Figure 3.1). The above procedures were repeated for English native speakers.

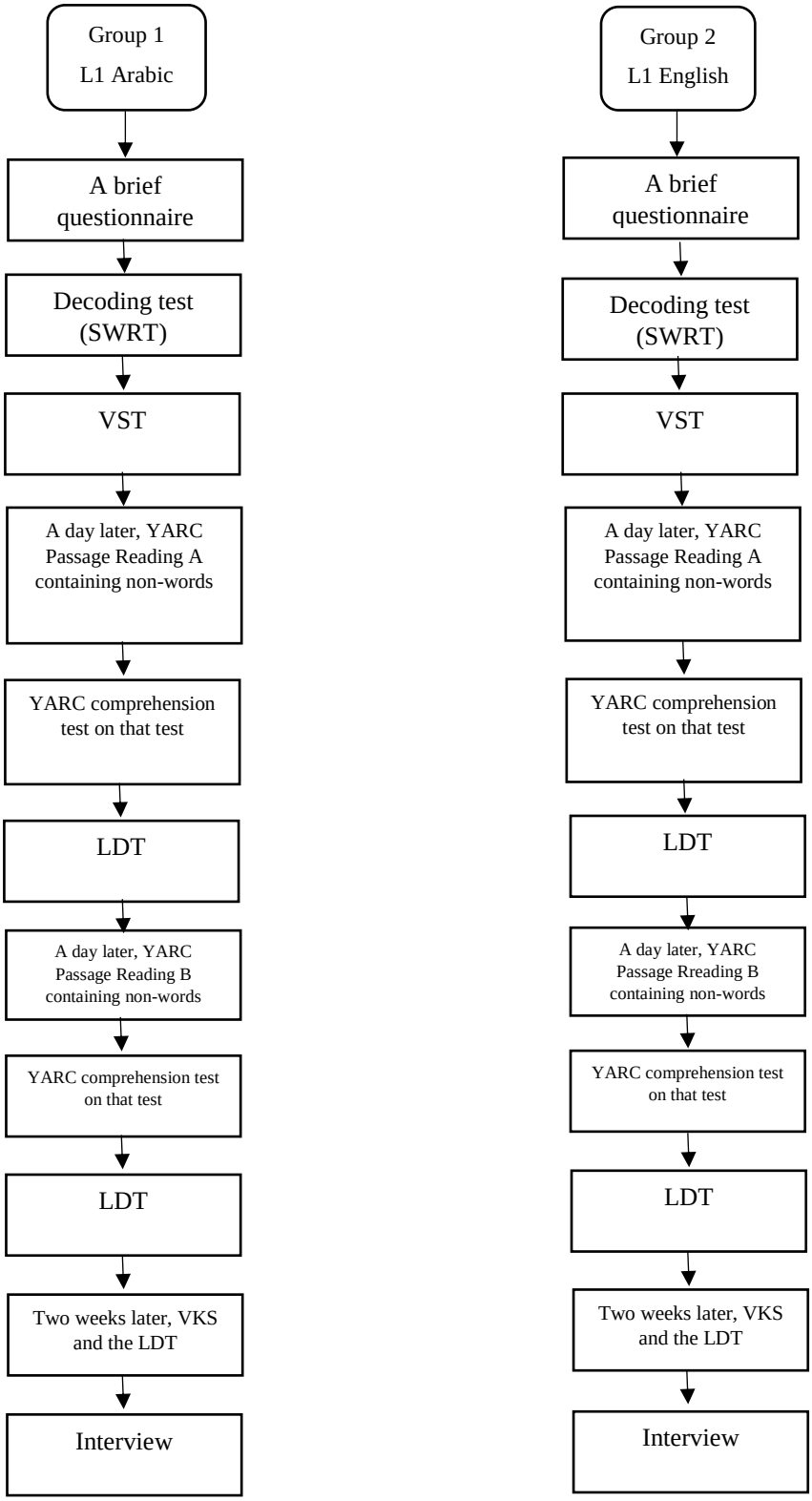


Figure 3.1 Procedure Study 1

3.3.11 Ethical approval procedure

Ethical approval for this study was gained through the standard procedure required by the Department of English Language and Applied Linguistics at the University of Reading (see Appendix 10).

3.4 Analysis of Results

3.4.1 Introduction

This section presents the results of Study 1. The results of all tests, the Vocabulary Size Test (VST), Lexical Decision Task (LDT), Vocabulary Knowledge Scale (VKS), and York Assessment Reading Comprehension Secondary Test (YARC), are presented, followed by the correlations and the analysis between the tests. A Kolmogorov-Smirnov test for normality was used to test all measures used in the study. Since the scores for all these tests, for both the English native speaker group (G1) and the learners of English (G2), were not normally distributed (see Appendix 11), non-parametric tests were used to investigate group differences. Two L1 Arabic learners of English from G2 were removed from the study because their accuracy scores fell below the criterion of 40%. In addition, four L1 Arabic learners of English in G2 were excluded from the LDT data set because their error rates exceeded 20%. As Jiang (2012) suggested, learners with high error rates (20% or higher) should be removed from the data set because measurements beyond the 20% cut-points are likely to skew the results of the data.

In this section, the Mann-Whitney U-test was used for between-group comparisons, whereas the Wilcoxon's Signed Ranks test was used for within-group comparisons. For each of these tests reported in this section, the value of $N_{(1)} = 30$ and $N_{(2)} = 28$, except in the LDT results, where the value of $N_{(1)} = 30$ and $N_{(2)} = 24$. An effect size was conducted to determine the level of each significant result reported in this section. According to Cohen's (1988) estimation of the effect size, d = 0.2 is a small, d = 0.5 is a medium, and d = 0.8 is a large effect size.

3.4.2 Comprehension scores

Table 3.8 presents the mean scores of three parts of the YARC Secondary Test: the Single Word Reading Test (SWRT), the reading comprehension questions (RCQ), and the summarisation of passages (RSQ). The mean accuracy score that G2 speakers obtained in the SWRT was 67.44% out of the maximum of 70 points. In the eight reading comprehension questions (RCQ), the mean accuracy scores for G1 participants were significantly lower than those obtained by G2 participants ($U = 258.5$, $p < 0.01$, d = 0.79). This indicates that the G2 group performed better in the comprehension task than the L1 English participants. As learners of English, it seems that they had received some language testing prior to the testing session, which may have helped them to perform better in the RCQ

than the G1 group. Lastly, no significant differences in accuracy were obtained between the two groups in the RSQ.

Table 3.8 Mean accuracy scores and standard deviations on the YARC Secondary Test

	G1 (N = 30)		G2 (N = 28)		Sig/Diff
	Mean	**SD**	**Mean**	**SD**	
SWRT (Max = 70)	_____	____	47.21 (67.44%)	2.97	___
RCQ (Max = 8)	5.43 (67.88%)	1.41	6.32 (79%)	1.02	0.01
RSQ (Max = 17)	11.90 (70%)	3.07	11.57 (68.06%)	2.77	NS

Performance on the eight reading comprehension questions was analysed as follows: by considering the scores from the four comprehension questions related to the target non-words, and those for the four comprehension questions relating to real words in the passages. Table 3.9 shows that both groups obtained similar mean accuracy scores in the comprehension questions related to the target words. By contrast, G1 obtained lower mean accuracy scores than G2 in the comprehension questions that do not relate to the target words; the difference in accuracy between the two groups is highly significant ($U = 168.00$, $p < 0.001$, d = 1.20). This means that, as observed above, G2 performed better in the RCQ (real words) than the G1 group. There was also no significant difference in accuracy between the RCQ related to the target non-words and the RCQ not related to the target non-words in the G1 group. However, within-participant comparisons revealed that the G2 group obtained lower scores in the RCQ related to the target non-words than those not related to the target non-words ($Z = -4.000$, $p < 0.001$, d = 2.31), whereas no significant difference between these conditions was observed for G1. This indicates that G2 faced more difficulties in understanding the comprehension questions containing the target non-words, while such a discrepancy in understanding was not found within the G1 group.

Table 3.9 Mean accuracy scores and standard deviations on the RCQ

	G1 (N = 30)		G2 (N = 28)		Sig/Diff
	Mean	**SD**	**Mean**	**SD**	
RCQ related to the target non-words (Max = 4)	2.67 (66.75%)	0.96	2.64 (66%)	0.78	NS
RCQ not related to the target non-words (Max = 4)	2.80 (70%)	0.76	3.68 (92%)	0.55	0.001

3.4.3 The association of reading comprehension with L2 readers' retention of the aspects of word knowledge (form, meaning, and use)

The first research question posed by this study was to investigate the aspects of word knowledge (i.e., form, meaning, and use) of new words that L2 readers encounter and retain in a reading comprehension task (YARC Secondary Test). As explained above, non-words were used as the target items for measuring the participants' ability to recognise the form, meaning, and use of newly learned items. Form recognition was tested with the LDT, and meaning and use were tested with the VKS test. As reported above, reading comprehension performance was measured via the YARC Secondary Test, which included reading comprehension questions (RCQ) and reading summarisation questions (RSQ) (see Instruments Section 3.3.2).

Descriptive statistics presented in Table 3.10 show that the G2 group generally obtained lower scores in the VKS for the target non-word category than in the LDT for the target non-word category. For the VKS measures, the overall mean was 4.88% (5.25% for providing the meaning, 4.5% for writing the non-words in a sentence). These component scores were considerably lower than those obtained for target non-words in the LDTs: 74% was obtained in the LDT (IPT) and 63.5% in the LDT (DPT). There was a highly significant difference in accuracy (Z = -4.183, p < 0.001, d = 2.58) between the VKS (meaning of the non-word) and the LDT (IPT) target non-word category for the G2, as well as between the VKS (using the non-word in sentences) and the LDT (IPT) target non-word category (Z = -4.172, p < 0.001 d = 2.56). There were also highly significant differences in accuracy between the VKS (meaning of the target non-word) and the LDT (DPT) target non-word category (Z = -4.199, p < 0.00, d = 2.61) and between VKS (using the target non-words in sentences) and LDT (DPT) target non-word category (Z = -4.188, p <0.001, d = 2.59) for the G2. This suggests that the G2 participants were better at retaining target non-word forms than word meanings and use.

Table 3.10 Mean VKS (target non-words), LDT (IPT and DPT) target non-words accuracy scores and standard deviations for learners of English (G2)

	G2		
	N	**Mean**	**SD**
VKS (target non-words) (Max = 8, including both meaning = 4 and using in sentences = 4)	8	0.39 (4.88%)	1.10
VKS (meaning of the target non-words) (Max = 4)	8	0.21 (5.25%)	0.57
VKS (using the target non-words in sentences) (Max = 4)	8	0.18 (4.5%)	0.55
LDT (IPT) (target non-words) (Max = 4)	4	2.96 (74%)	1.00
LDT (DPT) (target non-words) (Max = 4)	4	2.54 (63.5%)	1.06

The highest score obtained by only two G2 participants on the VKS was 4 out of the maximum of 8, whereas 24 (85.71%) obtained the lowest score of 0 on the same test. According to the LDT (IPT) results, only 1 of the participants obtained the lowest score of 0, and 8 achieved the maximum score of 4. In terms of the LDT (DPT), which was administered two weeks later, 4 of the G2 participants obtained the maximum score of 4, and only 1 participant obtained the lowest score of 0 (see Appendix 11b). Therefore, the G2 participants were better at retaining new word form than aspects of meaning and use.

In order to evaluate whether reading comprehension correlates with the acquisition and retention of the from, meaning and use of L2 word knowledge (RQ1), Spearman Correlations were computed to examine the relationship between the LDT (IPT), LDT (DPT), VKS, RCQ, and RSQ scores. A slightly stronger correlation was obtained between the VKS non-words category ($r = 0.496$, $p < 0.01$) and RSQ than between the VKS non-word category ($r = 0.474$, $p < 0.05$) and RCQ scores. Significant positive correlations were also obtained between the VKS and RSQ scores ($r = 0.465$, $p < 0.05$), and the VKS and RCQ scores ($r = 0.452$, $p < 0.05$) (see Table 3.11).

Table 3.11 Spearman Correlations between LDT (IPT), LDT (DPT), VKS, RCQ and RSQ for learners of English (G2)

	LDT (IPT, target non-words)	LDT (DPT, Target non-words)	LDT (IPT, total scores)	LDT (DPT, total scores)	VKS (target non-words)	VKS (total scores)	RCQ (total scores)	RSQ (total scores)
LDT (IPT, target non-words)	___							
LDT (DPT, (target non-words)	0.439*	___						
LDT (IPT, total scores)	0.316	-0.209	___					
LDT (DPT, total scores)	0.143	0.146	0.509*	___				
VKS (target non-words)	0.249	0.205	0.127	0.086	___			
VKS (total scores)	0.180	0.090	0.241	0.157	0.764**	___		
RCQ (total scores)	0.072	-0.067	0.166	0.170	0.474*	0.452*	___	
RSQ (total scores)	-0.110	0.119	-0.161	-0.016	0.496**	0.465*	0.611**	___

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

A regression model was also used to investigate the correlation between reading comprehension, and learning and retention of the three aspects of L2 word knowledge (form, meaning, and use). The independent variables were the reading comprehension scores (RCQ) and the reading summarisation scores (RSQ) for passages (A) and (B). The VKS, LDTs (IPT and DPT) target non-words scores were the dependent variable. Two separate regression analyses with a single predictor variable were computed, once with the RCQ and once with the RSQ.

With respect to the retention of meaning and use of the new words in a sentence, it was predicted (Prediction 1 of RQ1a) that higher comprehension scores of L2 learners of English (G2) will be associated with better retention of word meaning and use (in a sentence) on the VKS scores. The regression model suggested that reading comprehension, as measured with the RCQs of the YARC, did significantly correlate with the retention of meaning and aspects of the use of the target non-words ($\beta = 0.44$, $F = 6.39$ (1, 26), $p = 0.018$, $p < 0.05$). However, the model provided limited explanatory power, as suggested by an adjusted R^2 square value of 0.197. This means that 19.7% of the variance in the VKS target non-words scores was

accounted for by RCQ scores (see Appendix 11c). The regression model also suggested that the learners' ability to summarise stories, as measured with the RSQs, did significantly correlate with the retention of meaning and aspects of the use of the target non-words as measured with the VKS (B = 0.43, F = 6.05(1, 26), p = 0.021, p < 0.05). Again, the model provided relatively limited explanatory power (R^2 = 0.189). This means that 18.9% of the variance in VKS non-word scores was accounted for by the summarisation scores (see Appendix 11d).

With respect to the form of the new words (target non-words), it was predicted (Prediction 2 of RQ1a) that better comprehension (as measured with the YARC test including RCQ and RSQ) will be associated with less decay in form recognition of target non-words presented in the LDT across the two test periods. The results obtained in the LDT (IPT) suggested that reading comprehension, as measured with the RCQ and RSQ, did not correlate significantly with learning the form of the target non-words, (R^2 = 0.004) for the RCQ and (R^2 = 0.021) for the RSQ. In terms of learners' retention of the form of the target non-words in the LDT (DPT), the results from the regression model suggested that RCQ and RSQ scores did not correlate significantly with the retention of the form of the target non-words (R^2 = 0.021) for the RCQ scores and (R^2 = 0.019) for the RSQ scores. As can be observed from the mean accuracy scores of the LDTs (non-words category), shown in Table 3.10, the G2 group was also equally good at both LDTs, as there were no significant differences in accuracy between the two tests (immediate and delayed post-test).

To evaluate RQ1b, participants were interviewed after taking part in this study. They were asked how they determined the meaning of the four target non-words while reading the texts (see Appendix 5). According to their feedback, most of the participants reported that they were able to understand the meaning of these words. They explained that reading the text carefully helped them to guess the meaning of the target non-words. One of the English native speakers reported that she tried to understand and guess the meaning of the target non-words from the rest of the sentences in the stories. Another participant confirmed that the stories were coherent passages of logically connected sentences, so that he was able to provide the meanings of three of the target non-words correctly. However, one of the most common problems that the participants encountered, especially the L2 learners of English (G2), was that they showed limited vocabulary knowledge. For example, Student A (G2) said, "I cannot know some keywords [the target non-words] in the sentences, so that I struggled to understand the meaning of the sentence." Student B (G2) also

mentioned that "I have tried to know most of words, but I am still unsure about the meaning of these words." In addition, Student C (G2) said, "I suffer in reading. Some words are not easy to understand, I have never seen them before." Overall, based on their feedback, it seems that the G2 group faced more difficulties in understanding the target non-words than the G1 group, who were able to guess the meaning of these words easily.

In summary, the results fully supported hypothesis 1: The form of new L2 words were retained better than their meaning or aspects of use (in a sentence) (Jiang, 2000). Prediction 1 of the study was fully supported by the results, as learners who scored highly on a reading comprehension task also obtained high scores on the VKS. Prediction 2 was not supported by the results, in that reading comprehension did not correlate with the learning or retention of the form of the new words. It seems, therefore, that there is a positive association between reading comprehension and the retention of meaning and use of new words.

3.4.4 The contribution of vocabulary knowledge (size and depth) to reading comprehension performance

The second question of this study was to investigate how vocabulary size and vocabulary depth can explain variance in learners' reading comprehension performance. In other words, the objective was to determine whether a learner with a large vocabulary size would score better on the reading comprehension test (YARC Secondary Test), or whether it was a learner's depth of vocabulary knowledge that best explained this score. As already mentioned, vocabulary size was measured with the VST, and depth of word knowledge with the VKS scores (meaning and use in a sentence). Learners' reading comprehension performance was measured with the YARC Secondary Test, which included the RCQ and RSQ sections. The overall mean scores for VST measures for G1 and G2 are presented in Table 3.12. The maximum score for the first eight layers is 80. Table 3.12 shows that the G1 speakers obtained higher mean scores on the VST than G2. The difference in accuracy between the two groups was highly significant in the Mann-Whitney test U (U = -6.544, p < 0.001, d = 3.56).

Table 3.12 Mean accuracy scores and standard deviations for the VST

	G1		G2		Sig/Diff
	Mean	**SD**	**Mean**	**SD**	
VST (Max = 80)	73.20 (91.5%)	3.09	38.64 (48.3%)	11.74	0.001

The degree of difficulty of the VST was designed to gradually increase from words belonging to the most frequent layers to the least frequent layer. The results for each group, with respect to each frequency band, are presented in Figure 3.2. In this figure, the labels on the X-axis represent the words from the first 1,000 to the eighth 1,000 (Nation1 to Nation8), and the units on the Y-axis show the participants' scores from 0 to 10 in each 1,000 words (frequency band).

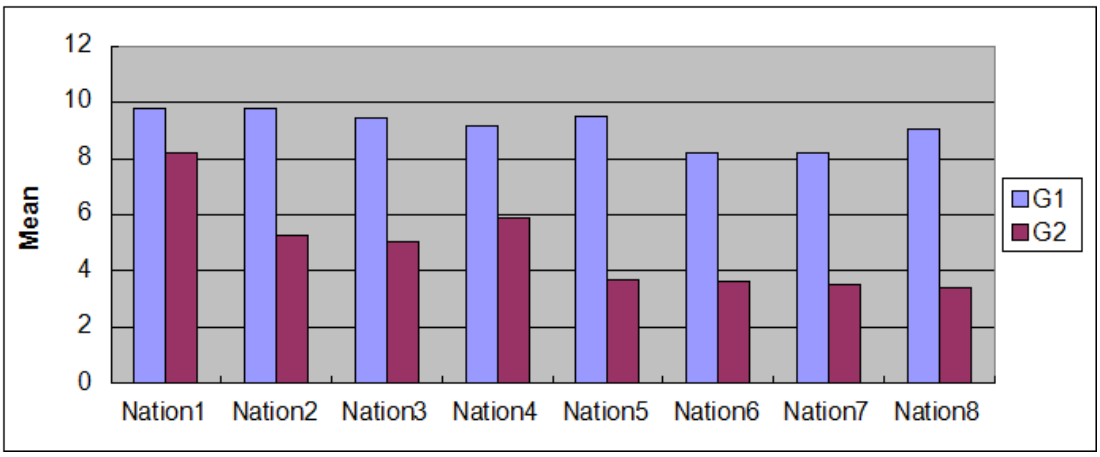


Figure 3.2 Nation's Vocabulary Size Test (Levels 1-8)

The results show that the G2 learners achieved the highest scores (10 points), for the first 1,000, and the lowest scores they obtained were in the last four 1,000 of the eighth 1,000. This means that they knew fewer words in the low frequency layers. Their responses for the bands 5,000 to 8,000 (Nation5 to Nation8) are below 50%, i.e., below chance level (see Figure 3.2). As shown in the repeated measures mixed model (ANOVA), the main effect of VST levels was highly significant where ($F = 46.37_{(7, 392)}$, $p < .001$ Eta Squared = 0.45). There was a significant interaction between VST levels and groups ($F = 18.53_{(7, 392)}$, $p < .001$; Eta Squared = 0.25). This means that the scores of the G2 group dropped faster than the G1 scores between VST levels. The main effect of groups was also significant ($F = 242.42_{(1, 56)}$, $p < .001$; Eta Squared = 0.81).

Table 3.13 presents the overall VKS results, and also those relating to real-word and target non-word categories. The mean accuracy scores of the VKS for both groups nearly reached the maximum of 12 points in the real-word category; however, participants obtained low scores in the VKS target non-word category.

Table 3.13 Mean accuracy scores and standard deviations for the VKS categories

	G1		G2		Sig/ Diff
	Mean	**SD**	**Mean**	**SD**	
VKS target non-words (Max = 8)	0.73 (9.17%)	1.60	0.39 (4.88%)	1.10	NS
VKS real words (Max = 12)	11.97 (99.17%)	0.18	11.75 (97.92%)	0.65	NS
VKS (meaning of the target non-words) (Max = 4)	0.47 (11.75%)	0.86	0.21 (5.25%)	0.57	NS
VKS (using the target non-words in sentences) (Max = 4)	0.30 (7.5%)	0.80	0.18 (4.5%)	0.55	NS

Spearman correlations between the VKS, VST, RCQ and RSQ scores were computed to evaluate the associations among these measures. It was predicted (Prediction 1 of RQ2a) that the association between learners' VKS scores and reading comprehension scores would be stronger than that between their VST scores and reading comprehension scores. The correlations in Table 3.14 show that this prediction was not borne out with respect to the RCQ, VKS and VST scores. However, it was confirmed with respect to (Prediction 2 of RQ2a) the association between learners' VKS scores and reading summarisation scores. As shown in Table 3.14, a moderately strong correlation was found between the VKS and RSQ scores (r = 0.397, p < 0.01), but the RSQ/VST correlation failed to reach significance. Another high correlation was found between RCQ and RSQ scores (r = 0.439, p < 0.01), which indicates that there is a positive association between comprehension scores and summarisation scores (see Table 3.14).

Table 3.14 Correlation between VST, VKS, RCQ and RSQ

	VST (total scores)	VKS (total scores)	RCQ (total scores)	RSQ (total sores)
VST (total scores)	___			
VKS (total scores)	0.366**	____		
RCQ (total scores)	- 0.119	0.111	____	
RSQ (total scores)	0.200	0.397**	0.439**	____

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

As shown in Table 3.15, a high correlation was obtained between the VKS (target non-words) category and the RSQ scores (r = 0.465, p < 0.01). Therefore, students who were good at summarising the stories were also good at retaining the meaning and use of new words.

Table 3.15 Correlation between VKS, RCQ, and RSQ

	VKS (target non-words)	VKS (real words)	RCQ	RSQ
VKS (target non-words)	___			
VKS (real words)	0.147	___		
RCQ	0.159	-0.032	___	
RSQ	0.465**	0.054	0.439**	___

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

The hypothesis relating to RQ2a was that vocabulary depth would explain additional variance in reading comprehension over and above variance explained by vocabulary size (Rashidi & Khosravi, 2010). A forced entry regression analysis was carried out with RCQ as the dependent variable, and the accuracy scores of the VST, and the VKS as predictors. The results from the multiple regression model suggested that VST, and VKS scores did not correlate significantly with reading comprehension scores.

Another regression analysis was carried out with RSQ as the dependent variable and the VST, and the VKS as predictor variables. Table 3.16 reveals that the multiple regression model results suggested that VKS scores significantly correlate with answering the summarisation questions of the reading comprehension section (β = 0.370, p = .006). The R square value was (0.160), which means that 16% of the variance in the RSQ scores was accounted for by the VKS scores. The results from the multiple regression model also suggested that VST scores did not correlate significantly with learners' RSQ scores.

Table 3.16 Correlation of the VST, and VKS with RSQ

Model	Unstandardised Coefficients		Standardised Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	2.255	2.978	_____	.757	.452
VST	.011	.020	.075	.577	.566
VKS	.712	.250	.370	2.852	.006

a. Dependent Variable: RSQ

With respect to RQ2b, it was predicted (Prediction 1 of RQ2b) that the patterns of performance measures will be generally similar across the L1 and L2 groups in terms of their word learning and retention from reading. As shown in Table 3.13, the results show that there

were no significant differences in accuracy between the two groups for both target non-word and real- word categories in the VKS. It will be remembered that the VKS scoring was based on how to use the word accurately, both from a semantic and a grammatical perspective. Both groups had low mean accuracy scores in both categories (meaning and use in a sentence), and there were no significant differences in accuracy between the two groups.

All participants received the LDT (IPT and DPT) in order to test their ability to recognise the form of new words. As mentioned in Section 3.3.6, the LDT stimuli in both passages consisted of four categories: 12 real words from texts A and B (category 1), 16 real words that did not occur in the texts (category 2), 4 target non-words (category 3), and 32 non-words that did not occur in the two texts (category 4). For each of the Mann-Whitney tests reported in the LDTs section, the value of $N_{(1)} = 30$ and $N_{(2)} = 24$. Tables 3.17 and 3.18 represent the percentage mean accuracy scores of real words (categories 1 and 2) and target non-words (category 3) for both group in the LDT (IPT and DPT). The differences in accuracy between real words and non-words are highly significant for the G1 in the LDT (IPT) ($Z = -3.744$, $p < 0.001$, d = 1.87) and in the DPT ($Z = -4.288$, $p < 0.001$, d = 2.52). For the G2, the differences in accuracy between real words and non-words are also highly significant in the LDT (IPT) ($Z = -3.589$, $p < 0.001$, d = 2.15) and in the DPT ($Z = -3.989$, $p < 0.001$, d= 2.81). It reveals, as expected, that the mean accuracy recognition of real words is higher than it is for non-words.

Table 3.17 Mean accuracy percentages and standard deviations on the LDT (IPT) target non-words and real words

	Target non-words (category 3) (Max = 4)		**Real words (category 1 & 2) (Max = 28)**		**Sig/Diff**
	Mean	**SD**	**Mean**	**SD**	
G1 N=30	67.50%	30.19	96.55%	10.42	0.001
G2 N=24	73.96%	24.98	96.43%	3.94	0.001

Table 3.18 Mean accuracy percentages and standard deviations on the LDT (DPT) target non-words and real words

	Target non-words (category 3) (Max = 4)		Real words (category 1 & 2) (Max = 28)		Sig/Diff
	Mean	SD	Mean	SD	
G1 N=30	51.67%	39.36	97.38%	3.74	0.001
G2 N=24	63.54%	26.56	97.47%	3.07	0.001

Table 3.19 shows the mean accuracy LDT (IPT) scores for the four categories of word stimuli for the two groups. It was found that recognition accuracy for both G1 and G2 was high in categories 1 and 2 (real words occurring in the texts and real words not occurring in the texts), and there were no significant differences in accuracy between the two groups. While G2 performed slightly more accurately than the G1 group on their recognition of non-words from the texts, the Mann Whitney U-test shows that there was no significant difference in accuracy between the two groups in category 3. The G1 speakers produced higher average accuracy scores for category 4 (non-words not occurring in the texts) than G2, which shows a highly significant difference between the two groups (U = 123.0, $p < 0.001$, d = 1.36).

Table 3.19 Mean accuracy scores and standard deviations for the LDT (IPT) stimulus categories

	G1 (N = 30)		G2 (N = 24)		Sig/ Diff
	Mean	SD	Mean	SD	
Real words from the texts (Max = 12)	11.70 (97.5%)	1.12	11.46 (95.5%)	0.83	NS
Real words not occurring in the texts (Max = 16)	15.33 (95.81%)	1.84	15.54 (97.13%)	0.66	NS
Target non-words (Max = 4)	2.70 (67.5%)	1.21	2.96 (74%)	1.00	NS
Non-words not occurring in the texts (Max = 32)	30.80 (96.25%)	1.63	27.50 (85.94%)	3.11	0.001

The delayed LDT (DPT) mean accuracy scores are presented in Table 3.20. The results reveal that both groups achieved form recognition of real words, whether they were included in the texts or not (Categories 1 and 2) and no significant differences in accuracy between the two groups were obtained. In category 3 (target non-words), again, no significant difference was obtained between the groups, but the level of performance for both groups was considerably

lower than was observed for real words. For both groups, correct rejection of non-words not occurring in the texts (category 4) was similar to that obtained in the immediate LDT test. The difference between the two groups was highly significant ($U = 85.50$, $p < 0.001$, $d = 1.71$) with G1 outperforming G2. Overall, it seems that the scores obtained by both G1 and G2 were quite similar to the immediate LDT post-test results. There were no significant differences between the two tests (immediate and delayed post-test) for both groups in a Wilcoxon's Signed Ranks test in categories 1, 2 and 4.

Table 3.20 Mean accuracy scores and standard deviations for the LDT (DPT) stimulus categories

	G1(N = 30)		G2 (N = 24)		Sig/Diff
	Mean	**SD**	**Mean**	**SD**	
Real words from the texts (Max = 12)	11.83 (98.58%)	0.38	11.75 (97.92%)	0.61	NS
Real words not occurring in the texts (Max = 16)	15.43 (96.44%)	0.94	15.54 (97.13%)	0.51	NS
Target non-words (Max = 4)	2.07 (51.75%)	1.57	2.54 (63.5%)	1.06	NS
Non-words not occurring in the texts (Max = 32)	31.17 (97.38%)	1.23	27.67 (86.47%)	3.12	0.001

In terms of form recognition of the target non-words (category 3), it can be observed that the G1 performance was significantly poorer at the LDT (DPT), where the mean score was 51.75% (Table 3.20), compared with the LDT (IPT), where the mean score was 67.5% (Table 3.19); ($Z = -2.264$, $p < 0.02$, $d = 0.91$). The G2 performed equally well on the immediate and delayed LDTs, as there was no significant difference between the two tests.

When calculating speed of processing means (RT), all error trials were excluded from the reaction time data. Tables 3.21 and 3.22 show the RT means of target non-words (category 3) and real words (Categories 1 & 2) from the LDT (IPT and DPT). The difference in mean RTs between real words and target non-words was significant for G1 using the Wilcoxon's Signed Ranks test for both the LDT (IPT) ($Z = -3.412$, $p < 0.01$, $d = 1.59$) and the DPT ($Z = -2.191$, $p < 0.05$, $d = 0.87$). Differences in mean RT for the G2 also reached statistical significance in the DPT ($Z = -2.889$, $p < 0.01$, $d = 1.46$). As expected, in all cases where a significant effect was observed, the RT mean scores of the real words were shorter than those for target non-words.

Table 3.21 Means and standard deviations of the RTs for target non-words and real words on the LDT (IPT)

	Target non-words (category 3)		Real words (categories 1 & 2)		Sig/Diff
	Mean	**SD**	**Mean**	**SD**	
G1 N=30	926.79	301.81	655.04	158.82	0.001
G2 N=24	1012.52	455.64	824.15	151.54	NS

Table 3.22 Means and standard deviations of the RTs for target non-words and real words on the LDT (DPT)

	Target non-words (category 3)		Real words (categories 1 &2)		Sig/Diff
	Mean	**SD**	**Mean**	**SD**	
G1 N=30	908.84	426.46	631.93	166.89	0.02
G2 N=24	1104.00	507.96	781.60	136.56	0.004

Tables 3.23 and 3.24 present RT means for the four categories of words in the LDTs: real words from the texts (category 1), real words not occurring in the texts (category 2), target non-words (category 3), and non-words not occurring in the two texts (category 4). The findings revealed that mean recognition response times were significantly shorter for L1 English native speakers (G1) than L2 learners of English (G2) in all categories except category 3, which contained the target non-words.

As shown in Tables 3.23 and 3.24, the differences between the two groups were highly significant in category 1 (real words from the text) in LDT (IPT) (U = 155.00, p < 0.01, d = 1.11) and in LDT (DPT) (U = 184.00, p < 0.01, d = 0.92). There were highly significant differences between the two groups in category 2 (real words not occurring in the text) in LDT (IPT) (U = 157.00, p <0.01, d = 1.10), and in LDT (DPT) (U = 164.00, p < 0.01, d = 1.05). There were also highly significant differences between the two groups in category 4 (non-words not occurring in the text) in the LDT (IPT) (U = 34.00, p <0.01, d = 2.43) as well as in the LDT (DPT) (U = 96.00, p <0.01, d = 1.60). Within-participant comparisons of processing speed in all four categories of words in the LDTs between immediate and delayed post-test revealed, no significant differences for the G1 group. There were also no significant

differences in response time between IPT and DPT for G2 for categories 1 and 2 (real words whether they occurred in the text or not). The difference between IPT and DPT just reached significance for the target non-words (category 3) (Z = -2.029, p = 0.04, d = 0.91), where response times were shorter in the immediate post-test. Conversely, mean response times in category 4 (non-words not occurring in the text) were shorter in the DPT than the IPT (Z = -2.971, p = 0.003 d = 1.53).

Table 3.23 Means and standard deviations of the RTs for LDT (IPT) stimulus categories

	G1 (N = 30)		G2 (N = 24)		Sig/ Diff
	Mean	**SD**	**Mean**	**SD**	
RT (real words from the texts)	647.49	156.91	837.88	196.74	0.001
RT (real words not occurring in the texts)	662.59	185.47	810.43	135.81	0.001
RT (target non-words)	926.79	301.81	1012.52	455.64	NS
RT (non-words not occurring in texts)	738.84	156.10	1229.29	241.27	0.001

Table 3.24 Means and standard deviations of the RTs for LDT (DPT) stimulus categories

	G1 (N = 30)		G2 (N = 24)		Sig/ Diff
	Mean	**SD**	**Mean**	**SD**	
RT (real words from the texts)	642.51	172.36	769.06	143.49	0.002
RT (real words not occurring in the texts)	621.35	187.80	794.15	164.52	0.001
RT (target non-words)	908.84	426.46	1104.00	507.96	NS
RT (non-words not occurring in texts)	742.40	195.27	1064.03	235.21	0.001

Therefore, in summarising the stories, the hypothesis for RQ2a was supported by the findings, since in the case of vocabulary depth, only the aspects of meaning and use explained additional variance in reading comprehension. Overall, Prediction 1 of RQ2a was not supported by the results, as there was no significant correlation between the VKS scores and reading comprehension scores. The results supported Prediction 2 of RQ2a, as the VKS obtains a high correlation with the RSQ scores. It was also found that students who were good at summarising the stories were also good at retaining the meaning and use of new words (see Table 3.15). For comparison between groups, the overall results show that both groups were equivalent in their learning and retention of new words and both groups did not

differ in their processing speed of the target non-words. This will be further explained in the discussion section.

3.4.5 The contribution of frequency to learning and retention the target non-words

The third research question of this study was to investigate whether frequency of occurrence of the target non-words or the frequency of the original words which were replaced by the target non-words explained learning and retention of these words. As described in Section 3.3.5.4, the four target non-words were controlled by their occurrence in the texts and categorised into two sets: two non-words, which occurred eight times (*cenedies* in passage A and *pocoko* in passage B), and two non-words which occurred four times (*toroko* in passage A and *sataca* in passage B). It was expected (RQ3 Prediction 1) that new words which occur more frequently in the text would be better learned and retained. Table 3.25 reveals the mean accuracy scores of the target non-words that occurred eight times and four times in the VKS and the LDT (IPT and DPT) for both G1 English native speakers and G2 learners of English. The Wilcoxon's Signed Ranks test results revealed that the within-participant difference for both groups just reached significance in the LDT accuracy scores (IPT): accuracy scores after eight repetitions were higher than after four repetitions for G1 ($Z = -2.399$, $p < 0.05$, $d = 0.97$), and G2 ($Z = -1.941$, $p < 0.05$, $d = 0.86$). The advantage of eight repetitions over four also just reached significance in the LDT accuracy scores (DPT), but only for G1 ($Z = -2.676$, $p < 0.05$, $d = 1.12$). No significant differences were found between these two repetition rates with respect to the VKS results for either group.

Table 3.25 Mean accuracy scores and standard deviations on the VKS, LDT (IPT and DPT) for target non-words appearing eight times or four times in the reading passages

		NW occurring 8 times			**NW occurring 4 times**		
		VKS (Max = 4)	**LDT (IPT) (Max = 2)**	**LDT(DPT) (Max = 2)**	**VKS (Max = 4)**	**LDT (IPT) (Max = 2)**	**LDT(DPT) (Max = 2)**
G1	Mean	0.47 23.5%	1.53 76.50%	1.23 61.50%	0.27 13.50%	1.17 58.50%	0.83 41.50%
	SD	1.14	0.68	0.90	0.64	0.75	0.83
G2	Mean	0.29 14.50%	1.63 81.50%	1.33 66.50%	0.17 8.50%	1.33 66.50%	1.21 60.50%
	SD	0.69	0.58	0.70	0.56	0.64	0.78

The response times (RTs) with which participants process target non-words appearing eight times and four times were also analysed (see Table 3.26). For both the IPT and the DPT, the Wilcoxon's Signed Ranks test revealed that for both groups there were no significant RT differences. This means that the frequency of occurrence of the target non-words in passages A and B did not significantly enhance the speed with which learners recognised the form of these words in the LDT both (IPT and DPT).

Table 3.26 Means and standard deviations of the RTs for target non-words appearing eight times or four times in the LDT (IPT and DPT)

	LDT (IPT)					**LDT (DPT)**				
	NW occurring 8 times		**NW occurring 4 times**		**Sig/ Diff**	**NW occurring 8 times**		**NW occurring 4 times**		**Sig/ Diff**
	Mean	**SD**	**Mean**	**SD**		**Mean**	**SD**	**Mean**	**SD**	
G1	897.19	401.11	979.34	516.95	NS	961.62	489.61	943.44	427.61	NS
G2	1122.52	478.62	944.22	618.29	NS	1207.63	601.60	938.51	493.19	NS

The potential contribution of the frequency of the original words (which had been replaced by target non-words) to learning and retention of these words was examined. Two target non-words replaced high frequency (1k) real words: *cenedies,* meaning "human" and *sataca*, meaning "trip," and two target non-words replaced low frequency real words: *toroko*, meaning "sting" (4k), and *pocoko,* meaning "suburb" (3k). It was predicted (RQ3 Prediction 2) that new words which replace high-frequency words in the text would be better learned and retained. Table 3.27 shows that there were no significant differences in accuracy between target non-words which replaced high frequency (HF) and low frequency (LF) words on the VKS for either group or the LDT accuracy scores (IPT and DPT) for the G1, which suggests that word frequency of the original words did not explain the learning and retention of the target non-words. For G2, by contrast, differences in LDT accuracy between these two frequency categories was obtained in the IPT (Z = -1.941, p < 0.05, d = 0.86) as well as in the DPT (Z = -2.595, p <0.01, d = 1.25). Contrary to expectations, however, the accuracy with which G2 recognised these words was higher for the LF items than for the HF items in the LDT (IPT and DPT).

Table 3.27 Mean accuracy scores and standard deviations on the VKS, LDT (IPT and DPT) for target non-words which replaced high-frequency words (HF) and low-frequency words (LF) in the reading passages

		HF – NW			LF - NW		
		VKS (Max = 4)	**LDT (IPT) (Max= 2)**	**LDT (DPT) (Max = 2)**	**VKS (Max= 4)**	**LDT (IPT) (Max = 2)**	**LDT (DPT) (Max = 2)**
G1	Mean	0.43 21.50%	1.37 68.50%	1.00 50 %	0.30 15%	1.33 66.50%	1.07 53.50%
	SD	0.94	0.67	0.91	0.70	0.71	0.78
G2	Mean	0.11 5.50%	1.33 66.50%	1.00 50 %	0.29 14.50%	1.63 81.50%	1.54 77 %
	SD	0.42	0.64	0.66	0.90	0.58	0.72

In terms of the RTs, as can be seen in Table 3.28, in the IPT, no significant differences were found for either group in RTs for items which replaced high-frequency (HF) or low-frequency words (LF). In the DPT, by contrast, a significant difference between these frequency categories was obtained (Z = -2.864, p =0.004, d = 1.44) for G2. The speed, however, with which respondents recognised these words, was shorter for the LF items than for the HF items in the LDT (DPT).

Table 3.28 Mean accuracy scores and standard deviations of the RT for target non-words which replaced high-frequency words (HF) or low frequency (LF) on the LDT (IPT and DPT)

	LDT (IPT)					**LDT (DPT)**				
	HF - NW		**LF- NW**		**Sig/ Diff**	**HF - NW**		**LF - NW**		**Sig/ Diff**
	Mean	**SD**	**Mean**	**SD**		**Mean**	**SD**	**Mean**	**SD**	
G1	955.54	495.0 3	894.1 7	428.3 1	N S	1001.0 8	536.2 2	910.0 7	684.8 2	NS
G2	1132.6 7	543.4 6	864.0 1	499.9 0	N S	1460.4 3	795.8 4	925.6 7	442.7 4	0.004

In summary, as expected, the LDT immediate post-test results of both groups, and the LDT delayed post-test results of only the G1 group, showed that target non-words which occurred eight times were better retained than target non-words which occurred only four times. It was also found that word frequency (high or low) of the words from the original text which were replaced by target non-words did not explain the learning and retention of these

target non-words. With respect to RTs, the results also revealed that frequency of occurrence of the target non-words and the high or low frequency of the original words which were replaced by the non-words did not significantly enhance the speed with which learners processed the form of the target non-words.

3.5 Discussion

3.5.1 Introduction

The purpose of this section is to discuss the results of Study 1. The following sub-section presents the key findings and discusses the theoretical implications in the light of the extant research presented in the literature review chapter and shows to what extent the current findings are consistent with those in the literature. The third sub-section offers some pedagogical implications of the findings that are relevant in L2 vocabulary teaching and learning.

3.5.2 Theoretical implications of results

The five key theoretical implications arising from the results presented in Study 1 are summarised in Figure (3.3).

Figure 3.3 Key findings of Study 1

This study's first objective was to investigate the aspects of word knowledge (i.e., form, meaning, and use) of new words that L2 readers encounter and retain in a designated reading comprehension task (the York Assessment of Reading for Comprehension Secondary Test). As expected and shown by comparing the LDT accuracy scores (IPT and DPT) with the VKS accuracy scores, the current results show that L2 learners of English are better at retaining word form than word meaning. The results, therefore, fully support Hypothesis 1: the form of new words will be retained better than their meaning or use. This study provides further evidence for Jiang's (2000) model of vocabulary learning among L2 learners, namely, that the form of an L2 lexical entry can be learned first, at an earlier stage.

As mentioned in the literature review, all word knowledge aspects are learned simultaneously, although some are assimilated sooner than others, due to the different stages of word acquisition (Schmitt, 1998). The current results also appear to support previous research conducted by Pigada and Schmitt (2006), which revealed that, even after a small number of exposures, spelling was strongly enhanced; meaning and grammatical aspects were also strengthened, although to a lesser degree. Similarly, a study by Brown et al. (2008) found that the more enhanced incidental vocabulary learning from reading was shown in the word form and meaning recognition, but far less with respect to producing the meaning. These findings, together with the current study's results, are consistent with Jiang's (2000) model, which proposed that at the first stage, the L2 lexical entry contains only the form (that is, the phonological form and information about the spelling). At this stage, the activation of the L1 lexical form is also possible, but with increasing exposure and the L2 learners' language proficiency, the second stage may be reached where L1 lemma mediation (conceptual representation) appears more directly, without relying as much on L1 lexical forms.

The current findings also suggest that the relationship between reading comprehension performance and retention of the meaning and use of L2 new words is positively related. As expected, L2 learners who score highly on reading comprehension tasks also obtain high scores on the VKS. Moreover, the findings show that L2 reading comprehension performance contributes significantly to retaining the meaning and use (in a sentence) of L2 new words. This seems to be consistent with Rashidi and Khosravi's (2010) view that depth of word knowledge has a positive significant correlation with reading comprehension in L2. However, no significant link was found among the L2 learners in the present study between reading comprehension performance and form recognition of new words. This illustrates that even when participants sufficiently learn and retain the forms of the new words from reading, this incidental receptive learning and retention may not be enhanced by comprehension performance. This may indicate that during reading, L2 learners focus on meaning and use more than on the form of words in order to interact with the texts for answering reading comprehension summarisation questions.

An interview was conducted to evaluate participants' familiarity with the target non-words. As shown by the results in section (3.4.3), it appears that the G2 group made more effort to understand the target non-words' meaning than the G1 group, who were able to guess the meaning of these words easily. Based on their feedback, most L2 learners revealed that they encountered challenges in identifying the correct meaning of target non-words from reading texts, whereas L1 speakers showed the possibility of guessing the meaning of the target non-

words from the rest of the sentences in the stories. This was confirmed by the reading comprehension results: it is shown that the G2 group performed better in the reading comprehension questions that do not relate to the target non-words than the G1 group, whereas no significant difference was found between them in answering the comprehension questions relevant to the target non-words. This clearly suggests that the G2 group's challenges posed by the target non-words may have led them to parallel the G1 group's performance in responding to the questions related to these non-words.

The study's second main aim was to find out to what extent vocabulary size and depth can explain variance in learners' reading comprehension performance. Previous research has found evidence that the size and depth of vocabulary knowledge have a distinctive relationship with reading comprehension. As shown in Chapter 2 (Section 2.5.1), vocabulary knowledge researchers have strongly emphasised the link between vocabulary size and success in reading comprehension (Laufer & Sim, 1985; Laufer, 1989; Hsueh-chao & Nation, 2000; Schmitt et al., 2011).

Conversely, the current findings revealed that when learners summarised stories, only the aspects of meaning and use (in a sentence) significantly explain variance (16%) in reading comprehension. This is an important finding that strongly suggests that those learners with greater awareness of these aspects are better at understanding words in depth and that both aspects can contribute positively to reading comprehension performance. This confirms a study conducted by Nergis (2013), who found that syntactic knowledge significantly predicts the reading comprehension in L2. Another study by Rashidi and Khosravi (2010) provides further evidence that depth of vocabulary is more important for reading comprehension than size of vocabulary. By controlling the size of vocabulary knowledge, they found that depth of vocabulary knowledge provides a stronger contribution for about 69% of unique variance to reading comprehension performance than the contribution provided by the vocabulary size (55%).

The current results show a significant positive correlation between depth of word knowledge and summarising the reading text. In other words, learners who are better at summarising stories are also proficient in retaining the meaning and use of new words. This, as expected, confirms that increasing the depth of vocabulary knowledge about particular words can contribute to successful reading. This finding concurs with earlier research that found that understanding words more deeply may lead to more flexible use, in which the words' semantic applications can be accessed more readily in various contexts (Anderson & Freebody, 1981).

This may then lead to improved accessibility, that could ultimately affect reading comprehension performance. Another important finding of Zahang and Yang (2016) was that vocabulary depth is the unique predictor of reading comprehension. For short passage comprehension particularly, vocabulary depth was a significantly stronger predictor of reading comprehension performance than vocabulary size. This unique contribution of vocabulary depth to reading comprehension suggests that the comprehension process demands more than only knowing the basic meaning of the words occurring in the text, but also other knowledge such as how those words relate to each other and how they can be used (Zahang &Yang, 2016). It thus seems reasonable to expect learners who have a greater depth of word knowledge to be more capable of determining the meaning of unknown words, which leads to success in reading comprehension.

The present results found no significant relationship between vocabulary size and reading comprehension performance. This contrasts with other findings (Liu & Nation, 1985; Laufer, 1992; Hsueh-chao & Nation, 2000) which suggest a high interrelation between vocabulary size and reading comprehension. However, this may have occurred because learners, particularly L2 learners – as in the present study – experience serious difficulties understanding reading comprehension questions. The current research showed that L2 learners have more difficulty understanding reading comprehension questions, especially when these are related to non-words. Moreover, learners' feedback revealed that they had trouble deriving the meaning of target non-words from reading texts and they showed more hesitation when identifying the correct meaning. Therefore, the results seem to show that L2 learners of English make a greater effort to answer reading comprehension questions than native English speakers do, however, they face difficulties in understanding and answering questions that specifically ask about target non-words in texts. This may indicate that because these reading comprehension questions involved knowledge of a word that they had recently learned, it may be very difficult for participants to use the words already in their vocabulary to contribute to their reading comprehension, and hence no relationship was found in the present study between vocabulary size and reading comprehension performance.

In terms of form recognition, percentage accuracy scores suggest that the participants in this study are better at retaining real words than target non-words. This finding was also confirmed by the RT scores for real words: response times were shorter for real words than non-words in both the immediate and delayed post-tests for the native English speakers, but only in the delayed post-test for L2 learners. This is to be expected, as learners are often more familiar with already known words than newly learned words.

The present study also demonstrated a decrease in the accuracy scores of target non-words among native English speakers from the immediate to the delayed post-test, which was expected because new words learned from reading are often forgotten. The L2 learners, interestingly, showed that retention of new words persisted for nearly two weeks, which confirms the findings of Horst et al. (1998) that retention can last for over ten days. The L2 learners' RTs for non-words that did not occur in texts were also shorter in the delayed post-test. Their RTs, however, for target non-words were shorter in the immediate than in the delayed post-test, which confirms the findings of a study conducted by Waring and Takaki (2003), which concluded that new words learned from reading can decline and may be completely forgotten within a few months.

The aim of RQ2b was to find out to whether the materials selected for the study were appropriate for L1 Arabic learners of English (G2). To achieve this, their performance was compared with that of L1 English native speakers (G1) in terms of word learning and retention from reading. As shown by the VKS and LDT results, both groups were equivalent in their learning and retention of new words. Both groups also did not differ in their processing speed for the target non-words. This clearly indicates that the task material selected for the present study were appropriate for L1 Arabic learners of English (G2). Since the L2 learner group in the present study acquired their L2 at an early age of about 9 to 10 years old, it is likely that early experience with the two languages (Arabic and English) may facilitate the L2 learners' ability to learn and retain L2 new words later on. Although most of the current L2 participants acquired their L2 through classroom exposure, linking L2 performance to early exposure to both languages can be reasonable. This is confirmed by previous research conducted by Kaushanskaya and Marian (2009), which has indicated that acquisition of any two languages early in life enhances the capacity to learn new words in adulthood. However, it is still unclear whether L2 performance can arise from early exposure to the two languages, or if any other factor yields this result. This explanation clearly demands further investigation into the mechanisms of bilingual performance compared to that of monolinguals.

A key finding of the current research was that frequency of occurrence positively enhanced the form recognition of target non-words for both groups in the immediate post-test, but only for native English speakers in the delayed post-test. These results concur with other findings on L2 learners that show they have a 5% chance of recognising a word's form or its meaning after at least eight repetitions to this word, which clearly shows the value of higher exposure frequency rates (Waring & Takaki, 2003). This finding also confirms earlier research conducted by Eckerth and Tavakoli (2012), revealing that, for L2 learners, the effects of frequent exposure

may disappear over time, as measured by delayed post-tests. Contrary to expectations, however, the cited study's RT results show that the frequency of occurrence of non-words does not explain the speed with which learners process the form of these non-words.

No significant differences were found between target non-words that replaced high frequency (HF) and low frequency (LF) words, indicating that the factor of word frequency of the original words does not noticeably increase learners' chances of learning and retaining the new words. This is because these original words did not appear in the texts, as they were replaced by the target non-words in the texts, and when participants read these target non-words in the texts, they may have mentally replaced those words with high or low-frequency words. As an example, the target non-word *cenedies* can mentally be replaced by the high-frequency item "human" or the low-frequency item "crowd." Therefore, this factor of word frequency of the original words which were replaced by target non-words does not seem to explain learning and retention of these new words. This is also confirmed by the RT results of LDT (IPT and DPT), as these reveal that the frequency of original words replaced by target non-words does not significantly explain the speed with which learners process the form of target non-words.

3.5.3 Pedagogical implications

The current findings show how new word acquisition can be enhanced by reading and to what extent vocabulary knowledge can affect reading comprehension performance, which can have positive implications for L2 vocabulary teaching and learning. Because some teachers tend to guide L2 learners directly in their learning process, an important pedagogical implication – particularly in the context of Saudi Arabia – is that instructors need to enhance the opportunities for incidental vocabulary learning from reading in L2 classrooms. This can increase L2 learners' motivation (Akasha, 2013) and aptitude for acquiring new lexical items. Teachers should facilitate incidental exposure to new words while reading by encouraging their students to read in order to understand, rather than directing them in any conscious way to learn new words.

Furthermore, as Nation (2001, p. 238) points out, "learning rates can be increased considerably by paying some deliberate attention to vocabulary." Thus, teachers need to be aware of the importance of vocabulary activities designed to increase learners' vocabulary size. Engku Ibrahim, Sarudin and Muhamad (2016) strongly recommend that instructors keep increasing learners' knowledge of HF words, and, simultaneously, expand their current knowledge of LF words. Teachers should also encourage their students to engage in a variety

of extensive reading activities, as multiple studies have shown the importance of reading in enhancing learners' vocabulary size and reading capacities (Krashen, 2004; Nation, 2001).

In addition, teachers need to be careful of any integration of these specific aspects that may occur either in the L1 or L2. Instructors must also be sure that learners receive adequate contextualised input by linking L2 words with the L2 context in order to minimise fossilisation of L2 words. As Jiang (2000) explains, a fossilised feature of L2 words is that L2 lexical entries are represented without a fully developed L2 lemma, and a weak link is made between the concepts in question and L2 words.

As the present study reported, the significant role of vocabulary depth in reading comprehension suggests that depth of word knowledge should be considered an important factor in reading comprehension. Teaching in depth new L2 vocabulary in texts, such as exposing students to multiple word meanings and relationships, can facilitate the comprehension of those texts. Another important implication for teachers is the need to be careful in developing vocabulary learning tasks. To provide better conditions for developing L2 lexical items, instructors, therefore, need to focus separately on each lexical aspect, including, among others, formal, semantic, syntactic and morphological features, when teaching new lexical items.

As shown in the results of the present study, learners are likely to forget newly learned words after a few days. Teachers, therefore, need to concentrate on how well new words are retained, for example, by providing practice sessions that may help learners process new words more deeply and maintain their initial level of learned vocabulary. The present research, in addition, confirms the importance of the frequency of exposure factor. Thus, as shown in earlier studies (Saragi et al., 1978; Waring & Takaki, 2003), teachers are strongly recommended to provide regular repeated exposure to new words. For instance, instructors can develop reading tasks in which students can encounter the same new words several times with different meanings in the same text, which could significantly affect their ability to learn these words. Overall, given the key practical implications of the present study highlighted above, L2 reading classes should address two essential elements: the importance of increasing the size of L2 learners' vocabulary, and ways to expand the depth of their knowledge of these words.

The second part of the current study is an intervention study, designed after obtaining the results of Study 1, in order to examine the role of a post-reading task with a high involvement load in enhancing L2 learning and retention of the meaning of new words encountered through

reading. Study 2, then, is based on Laufer and Hulstijn's (2001) Involvement Load Hypothesis (ILH) and aims to contribute to the discussion by examining to what extent a high involvement component (depth of processing levels) can contribute retention of the meaning of new words encountered in a reading task. Study 2 will be further discussed in the next chapter.

Chapter 4: Study2

4.1 Aims of the Study

Study 2 focuses on the correlation between depth of processing of new words in a task among L2 learners of English and learning and retention of the meaning of these words by these learners. Even with the contributions provided to support the claim made by the Involvement Load Hypothesis (ILH), there are many questions that yet need to be addressed. There is some evidence that tasks with a high involvement load, in particular, those with a strong *Evaluation* component, are associated with better learning and retention of new words (Keating, 2008; Kim, 2008). It is, hence, important to investigate the relationship between administering a task which requires learners to use a high level of *Evaluation* and L2 retention of the meaning of new words encountered in a reading task. For this study, a new Elaboration Task (ET) was developed as a post-reading activity to examine the association of different levels of depth of processing with learning and retention of the meaning of new words.

Based on Jiang's (2000) model, the current study further investigates how cross-language semantic priming can enhance word recognition in L2 learners of English. It aims to determine connections between L2 words and conceptual representation, i.e., whether participants continue to map the L1 lemma content onto the target L2 form or whether they have started to build lemmas with L2 word meanings. The study also aims to examine whether the group which had carried out an elaboration task with these target non-words were better at gaining the L2 lemma than those who had not carried out this task with these non-words.

In the light of ILH, there is still a question that needs to be answered as to whether the task itself is the only important factor or whether the number of exposure times the learner needs in order to learn the target words is also relevant (Hulstijn & Laufer, 2001). Study 2, then, also aims to investigate whether the frequency of occurrence of target non-words in a text (i.e., target words appeared either four or eight times in the passages), or the frequency band (high or low) of the original words which were replaced by the target non-words can explain learning and retention of these words.

4.2 The Research Question and Predictions

Based on Study 1, this investigation sought to answer the following research question:

RQ1: To what extent can an elaboration task which requires respondents to evaluate the meaning of a set of non-words contribute to retention of the meaning of these non-words that had been encountered in a reading task beforehand?

Hypothesis: An elaboration task which requires learners to use a high level of *Evaluation* will enhance L2 retention of the meaning of new words encountered in a reading task (Hulstijn & Laufer, 2001).

Prediction 1: The retention of the meaning of target non-words will be better for a group which had carried out an elaboration task with these non-words than for a control group. This will be the case in offline as well as online post-tests.

RQ2: To what extent does cross-language semantic priming enhance word recognition in L2 learners of English?

Hypothesis: Obtaining a priming effect from an L2 semantically related prime word will indicate a strong connection between L2 words and concepts (Jiang, 2000).

Prediction 1: L2 learners will succeed in obtaining a priming effect from an L2 semantically related prime word if that word has already reached the L2 integration stage (Jiang, 2000). For words which are still at the L1 mediation stage, the L2 prime which is semantically related to the L1 target word will not facilitate word recognition.

Prediction 2: L2 learners who have carried out an elaboration task with target non-words will succeed in obtaining a priming effect from an L2 semantically related prime word.

RQ3: To what extent can frequency of occurrence and frequency bands of target words explain learning and retention of these words encountered through reading?

Prediction 1: New words which occur more frequently in the text will be better learned and retained.

Prediction 2: New words which replace high-frequency words in the text will be better learned and retained.

4.3 Methods

4.3.1 Participants

L1 Arabic learners of English from undergraduate classes at the University of Reading were asked to participate in the present study. There was a total of 40 participants (28 females and 12 males) from different countries. 35 participants were from Saudi Arabia, and the remaining ones were a Libyan, a Yemeni, a Qatari, a Jordanian, and an Iraqi. Participants' ages ranged from 25 to 35 years, with a mean age of 28.7 years. All of them had been exposed to English since they were 12 years of age. Their exposure to English had been predominantly in formal instructional settings and had lasted a minimum of 13 years for the youngest participant in the group. Based on their IELTS scores, their English proficiency ranged from 5.5 - 6.0,

which translates into the B2 upper intermediate level, according to the Education First Standard English Test (EF SET) (see Appendix 18). These participants' reading proficiency in Arabic is the same as that reported for the Arabic L1 participants in Chapter 3 under subheading 3.3.1 (Qassim University Participants).

4.3.2 Research instruments

An offline word knowledge test was used in this study to examine the participants' depth of vocabulary knowledge: the Vocabulary Knowledge Scale (VKS) (Brown, 2008) (see Appendix 12) was modified from Wesche and Paribakht (1996). A Semantic Priming Task (SPT), based on Forster and Davis (1984), was designed for the purpose of the present study to assess online learning and retention of the meaning of the target words, and was conducted in two test sessions: immediate post-test (IPT) and delayed post-test (DPT), administered one week later. The York Assessment of Reading for Comprehension (YARC) Secondary Test, designed by Stothard et al. (2010), was used to test the participants' reading comprehension performance (see Appendixes 13a – 13d). An Elaboration Task was developed for this study as a post-reading activity to examine how different levels of depth processing affect retention of the meaning of new words (see Appendixes 14a – 14d). A Digit Span Test was used at the end of the study to measure the participants' working memory capacity (see Appendixes 15a – 15b). Each of these instruments are presented in more detail from Section 4.3.4 to Section 4.3.8.

4.3.3 Overall design

The present study was developed as a cross-sectional study in that it involved a comparison between groups. The participants were divided into two groups: one group took an elaboration task that focused on the meaning of non-words (ETA1-B1), and the other group carried out an elaboration task that did not focus on the non-words (ETA2-B2). All 40 participants took the ETs, but the order of presentation was reversed; this means that the group which had received A1 took also B2 and the group which had taken A2 received B1, in order to avoid the repetition effects. Hence, the number of L2 participants in each group was 40. The study also has a correlational component, as it investigated the relationship between the involvement load of a task (depth of processing levels) and the retention of the meaning of new words. A detailed description of the dependent and independent variables will be given in the following sub-sections.

4.3.3.1 Variables relating to RQ1

The first independent variable was the grouping variable involving the Elaboration Task (ET with or without target non-words). The dependent variables were the target non-word scores of the VKS and SPT (IPT and DPT, including accuracy target non-word scores and RTs) (see Table 4.1).

4.3.3.2 Variables relating to RQ2

The independent variable was the priming condition in the SPT (IPT and DPT). There were two conditions for each target non-word: Arabic root prime (1), Arabic and English prime (2). The dependent variables were the target non-word accuracy scores and RTs on the SPT (IPT and DPT) (see Table 4.1).

4.3.3.3 Variables relating to RQ3

Frequency of occurrence of non-words in the texts (four versus eight occurrences) and the frequency of original items which were replaced by the non-words in the texts (high versus low) were the independent variables. Scores on the VKS, and SPT (IPT and DPT, including accuracy scores and RTs), were the dependent variables (see Table 4.1).

Table 4.1 The overall design of Study 2

RQ	Prediction	Group	Independent Variables	Dependent Variables
1	1	ET (A1-B1) group and ET (A2-B2) group were compared	Grouping variable involving the Elaboration Task (ET with or without target non-words)	VKS scores on target non-words and accuracy scores and RTs for target non-words on the SPT (IPT and DPT)
2	1	ET (A1-B1) group and ET (A2-B2) group were combined	Priming conditions in the SPT (Arabic root prime versus Arabic and English root prime)	Accuracy scores and RTs for target non-words on the SPT (IPT and DPT)
2	2	ET (A1-B1) group and ET (A2-B2) group were compared	Priming conditions in the SPT (Arabic root prime versus Arabic and English root prime) and the grouping variable involving the Elaboration Task (ET with or without target non-words)	Accuracy scores and RTs for target non-words on the SPT (IPT and DPT)
3	1	ET (A1-B1) group and ET (A2-B2) group were combined	Frequency of occurrence of non-words in the text (four versus eight occurrences)	VKS scores on target non-words and accuracy scores and RTs for target non-words on the SPT (IPT and DPT)
3	2	ET (A1-B1) group and ET (A2-B2) group were combined	Frequency of original items in the text (high versus low)	VKS scores on target non-words and accuracy scores and RTs for target non-words on the SPT (IPT and DPT)

4.3.4 The York Assessment of Reading for Comprehension

As discussed in Chapter 3 (section 3.3.5.3), the YARC Secondary Test was originally designed to measure the reading comprehension of English-speaking children who are studying in secondary school and aged from 11 to 16 years old. The manual Stothard et al. (2010, p. 114) reveals, however, that children for whom English was an additional language were also included in the sample on which the norms were based. Stothard et al., (2010), report that the SWRT scores provide a reading age score for children whose first language is English (see Appendix 4b). In terms of L2 participants' SWRT scores, L1 Arabic adult students who participated in the current study have an average reading age which is equivalent to a native speaker age of 11 years and three months (see Table 4.2 and Appendix 4b). Since the SWRT scores for L1 Arabic adult learners fell within the range for native English children, the YARC Secondary Test was deemed suitable for the current study with L1 Arabic learners of English.

Table 4.2 Mean accuracy scores and standard deviations of L1 Arabic students (G2) on the SWRT and age score, based on norms from children whose first language is English

SWRT (scores)		**SWRT**
Mean	**SD**	**(year)**
47.98 (68.54%)	2.50	11.03

Since there is dearth of tests that are specifically developed for bilinguals and L2 learners, it is important to find out whether existing tests such as the YARC are suitable for non-native speakers. As mentioned in Chapter 3 (Section 3.3.5.3), one of the first to apply the test to adult L2 learners was Huang (2013) who tried this test out with adult Chinese L2 learners of English and concluded that the supplementary YARC passage can be more appropriate for this group of learners. The current study built on this first attempt and would enhance in evaluating YARC suitability for L1 Arabic adult learners of English. As mentioned in Chapter 3 (section section 3.3.5.3), it included two sections: decoding, which measures reading accuracy through the SWRT, and reading comprehension (literal and inferential meaning). These two components are addressed in the following sections.

4.3.4.1 Single Word Reading Test (SWRT)

The decoding test (SWRT) was administered in the present study for just a few minutes. Its function and its scoring procedures are the same as that reported for Study 1 (see Chapter 3, Section 3.3.5.2).

4.3.4.2 Passage Reading (A) and (B), including the target words (non-words)

The YARC passage reading is one part of the YARC assessments to measure reading comprehension. It comprises two sets of passages for each level: fiction (A) and non-fiction (B). In the present study, based on learners' raw scores (which ranged from 43-51) on the SWRT, the same two passages used in Study 1, Level 1 of the YARC Passage Reading test, were selected. The same two passages were also chosen for all participants to make sure that all participants were exposed to the same target words.

The target words in passages A and B used in the present study were the same as those in Study 1; 12 from the original text of approximately 480 words, so that 97.5% of the words of each text were already known and 2.5% of the text was composed of target non-words. As mentioned in Chapter 2 (Section 2.9), Hulstijn and Laufer (2001) did not use a pre-test to assess the participants' familiarity with the target words, and prior knowledge was tested by a post-exposure questionnaire. Also, as clarified by Martínez-Fernández (2008), the word types in Hulstijn and Laufer's (2001) study were not controlled. The target words used in their study consisted of expressions and different classes of words. The current study attempted to overcome all these issues by using non-words which had never been seen by the participants before, and all non-words replaced nouns in the texts, so that the word class was kept constant. These were the target words the participants were assumed to learn incidentally from reading. It is also important to know whether the task itself is the only essential factor, or whether the number of exposure times the learner needs to learn the target words is also important (Hulstijn & Laufer, 2001). Hence, the target words were categorised into two sets: non-words appearing eight times or four times, and non-words which replaced high frequency (HF) or low frequency (LF) words (see Table 4.3).

Table 4.3 The categorisation of the target non-words in the texts (A & B)

Passage A			Passage B		
Target words	**Frequency level**	**No. of occurrences in the text**	**Target words**	**Frequency level**	**No. of occurrences in the text**
cenedies "human"	1k	8 times	*sataca* "trip"	1k	4 times
toroko "sting"	4k	4 times	*pocoko* "suburb"	3k	8 times

There were two multiple-choice comprehension questions for passage (A) and two multiple-choice comprehension questions for passage (B). The reading comprehension questions were manipulated to include a reference to the target non-words and were selected from 13 comprehension questions for each passage, on the basis of the most relevant to the target non-words. The multiple-choice comprehension questions were designed to be relevant to the target non-words, so the participants needed to know the meaning of the non-words in order to complete the task, and the factor *Need* was clearly present in a moderate level, because it was externally induced, i.e., by the task itself. Each correct answer of the multiple-choice reading comprehension questions was worth one point (see Appendixes 13b and 13d).

The target non-words were highlighted in bold print in order to help the participants notice the words (Schmidt, 1994) and were glossed at the end of the two passages alphabetically, so the factor *Search* was omitted. *Search* was found to have no impact on L2 incidental vocabulary acquisition (Tang & Treffers-Daller, 2016); for this reason, the *Search* component was omitted in the current study. The meaning of the target non-words was provided in Arabic (the participants' L1) and was offered with three other incorrect meanings to encourage the participants to compare them and to choose the most suitable one for the given context (see Appendixes 13a and 13c). The purpose of the glosses was to encourage learners to focus on the meaning aspect of the words. As mentioned in the previous paragraph, the multiple-choice reading comprehension questions also focused on the target non-words. Therefore, the participants were expected to know the correct meaning of the target non-words if they had understood the text, and then to be able to answer the multiple comprehension questions correctly. If they had misunderstood the meaning of the target non-words, this would have affected their performance and error rates in multiple-choice comprehension questions, VKS and SPT. Research shows that tasks with a stronger *Evaluation* index can lead to processing a word with greater involvement, not only than tasks with a moderate *Evaluation* level, but also than tasks which involve the other two components (Kim, 2008). There was a second task, an elaboration task (ET), with a strong *Evaluation* component, in which students were required to

evaluate the meaning of the target non-words encountered through reading. In this way, a strong *Evaluation* was also triggered. Further details about the ET will be given in the next section.

4.3.5 Elaboration task

An elaboration task consisting of 32 sentences was developed as a post-reading activity to help participants reflect on the meaning of the target non-words presented in the stories. This would help to examine the role of depth of processing levels in the retention of the meaning of new words. There were two Elaboration Tasks (ET): an ET which focused on target non-words from Passage A (A1) and Passage B (B1), and an Elaboration distractor task was used for both passages (A2 and B2) which did not contain target words (see Appendixes 14a – 14d). The order of presentation was counterbalanced; this means that the group who had received the ET A1 took the B2, and the group who took the A2 received the B1, in order to avoid order effects. Each ET that included target non-words (A1 and B1) was split into two categories: 16 sentences containing target non-words and 16 containing real words. Each of these two categories were also divided into three types of sentences: grammatically and semantically appropriate sentences with non-words, grammatically and semantically appropriate sentences with real words, semantically anomalous target sentences with non-words, semantically anomalous sentences with real words, grammatically incorrect target sentences with non-words, and grammatically incorrect sentences with real words (see Table 4.4). The grammatically incorrect sentences were added to be able to compare students' ability to extract ungrammaticality and semantically anomalous sentences. Semantically anomalous sentences with real words were also included to compare students' ability to deal with semantically anomalous real words and semantically anomalous target non-words, such as *cenedies*. Participants were asked at the beginning of the task to consider the target non-words as normal English words. They were instructed that "Some of the following sentences contain words you have learned in the reading task. In doing the task below, please consider them as normal English words." Then, they were requested to judge how appropriate these sentences were on a six-point scale (1 = completely inappropriate and 6 = completely appropriate). They were also asked to provide an explanation by writing a sentence to support their judgment. Three example sentences were given at the beginning of the task, including all three types of sentences: grammatically and semantically appropriate, semantically anomalous, and grammatically incorrect, in order to explain clearly how the anomalous and ungrammatical sentence types were inappropriate (see Appendixes 14a – 14d).

Table 4.4 The number of items in each of the ETs that included target non-words (A1 and B1)

Sentence types	Real words	Target non-words
Grammatically and semantically appropriate	8	8
Semantically anomalous	4	4
Grammatically incorrect	4	4
Total	16	16

The ET-distractor samples (A2 and B2) included only real words, and the sentences were also divided into three types: grammatically and semantically appropriate sentences with real words, semantically anomalous sentences with real words, and grammatically incorrect sentences with real words (see Table 4.5).

Table 4.5 The number of items in each of the ETs A2 and B2

Sentence types	Real words
Grammatically and semantically appropriate	16
Semantically anomalous	8
Grammatically incorrect	8
Total	32

4.3.6 Vocabulary Knowledge Scale

A simplified updated version of the VKS developed by Brown (2008) was used in the present study, as it consists of two independent tests of word knowledge: receptive and productive**.** The VKS comprises a four-point elicitation scale to capture specific stages in the initial development of core knowledge (see Table 4.6). The unknown word category includes unknown words (Category I) and partial knowledge (form) (Category II). The known word category includes receptive word knowledge (meaning) (Category III) and productive word knowledge (use) (Category IV). The scale ranges from I, which represents total unfamiliarity, to IV, which represents the capacity to use a word grammatically and semantically accurately in a sentence. In the present study, the participants took the VKS of two target non-words from passage (A) combined with four real words and another VKS of Passage B, including the two target non-words from (B) combined with four real words, to assess the meaning and use aspects of those words retained over time (see Appendix 12).

Table 4.6 Modified Vocabulary Knowledge Scale

	Self-Report Category
I.	I don't remember having seen this word before.
II.	I've seen this word/phrase before, but I don't know what it means.
III.	I know what this word. It means _____ (Give the meaning in English or Arabic.)
IV.	I can use this word in a sentence _____ (Write a sentence.) (If you do this section, please also complete III.)

There were no points given to the unknown word categories, Category I, "I don't remember having seen this word before." Because the current study focused on the aspect of meaning of word knowledge, no points were also given to Category II, "I have seen this word before but I don't know what it means." That is, if learners chose Category I or II, they did not gain any points. The participants were awarded two points in the known word category: one for written receptive word knowledge if a synonym or translation of the target word was accurate. A wrong synonym or translation response was, however, given zero points. Learners were also awarded one point for productive word knowledge when the use of a target word was both grammatically and semantically accurate.

4.3.7 Semantic priming task

Priming tasks can provide data on aspects of lexical processing in reading, minimising the involvement of strategic effects or intentional processes because the respondents do not consciously perceive the prime (Jiang, 2000; McRae et al., 2005). For this study a semantic priming task was developed to investigate whether the new words were at the L1 lemma mediation stage or at the L2 integration stage (Jiang, 2000). In order to test Jiang's (2000) model, a masked cross-language priming experiment was used to examine connections between L2 words and conceptual representation. As an example, for the target non-word *toroko,* which was given the meaning "sting," the semantically related word via the Arabic root prime (condition 1) was "bite," and the semantically related in English and Arabic prime (condition 2) was "bee." The assumption was that learners would be faster at recognising new L2 words if the prime consists of a word that is semantically related to the target in the L1 only (and not in the L2). If this occurs, it can be assumed that the word is still at the L1 lemma mediation stage. If, however, the word was already at the L2 integration stage, learners would be faster at recognising a target word that is primed by a word that is semantically related in either the L1 or the L2.

4.3.7.1 Administering and scoring the test

In Semantic Priming Tasks (SPT) (Forster & Davis, 1984), participants are shown a mask (#####), then a prime, before a target word on which they are asked to make a semantic decision. In the current study, the semantic relation between prime and target words was manipulated to test the potential effect of certain variables on the participants' response. The masked priming task is very complex: it accesses various levels of cognition, depending on the length of the prime exposure, often called Stimulus Onset Asynchrony (SOA). It appears that very short SOAs, i.e., around 30ms, access morphological but not semantic processing. For semantic processing, the SOA should be around 80ms, but partial processing can occur at 60ms (Silva & Clahsen, 2008). For example, Pliatsikas and Marinis (2013) used an SOA of 33ms because they were focusing on morphological processing only. It was therefore necessary in the current experiment to use 60ms for the prime exposure of the semantic processing in order to reduce possible effects of memorising the words presented, or the opportunity for participants to develop response strategies, such as recognising that primes and targets often share common features (Rastle, Davis, Marslen-Wilson & Tyler, 2000). Silva and Clahsen (2008) proposed using 60ms for the prime exposure, to encourage partial semantic processing to take place, since a longer (e.g., 80ms) SOA might result in the presentation of the prime becoming visually detectable, thus causing a potentially distracting effect. The length of the target time cut-off was 2000ms (2 seconds) in order to avoid participants feeling too pressurised. Therefore, the timing sequence used in the current experiment was as follows:

a) mask: 500 ms

b) prime: 60 ms

c) target: 2000 ms.

Four types of primes were used with each target non-word: (1) semantically related to the target non-word via Arabic only, but not semantically related in English, (2) semantically related to the target non-word in English and Arabic, (3) semantically unrelated real words, and (4) non-words not seen by participants before (see Tables 4.7, 4.8, and 4.9). It is possible to determine whether participants have started to build lemmas with L2 word meanings or whether they continued to map the L1 lemma content onto the target L2 form. If participants had started to build L2 lemmas, accuracy scores in the SPT (IPT and DPT) would be ranked as follows: 2 > 1 and the reaction time (RT) would be 2 < 1. If they were still relying on the L1 semantic

representation of the target, then the rank order of the accuracy scores would be 1 > 2, and that of the RT would be:1 < 2.

The masked priming task was run four times for each reading passage, as four lists in which the priming type 1, 2, 3 or 4 only occurred once in each list. This meant that each participant performed the masked priming task four times, once for each of the four lists. As there were two target non-words in each list, different conditions (1, 2, 3, and 4) were used in the same list.

Table 4.7 The list of priming conditions used with each target non-word

Primes	Targets
Semantically related real words via the Arabic root	Target non-words from the reading tasks
Semantically related real words in English and Arabic	
Semantically unrelated real words	
Non-words not encountered in any reading task	

Table 4.8 The masked priming lists for passage A

List	Priming conditions	Targets	
		Prime	Target/Meaning
List 1	Semantically related via the Arabic root	personality	*cenedies*/human
	Semantically unrelated real word	product	*toroko*/sting
List 2	Semantically related in English and Arabic	human	*cenedies*/human
	Non-word	reseri	*toroko*/sting
List 3	Semantically unrelated real word	forest	*cenedies*/human
	Semantically related via the Arabic root	bite	*toroko*/sting
List 4	Non-word	mofata	*cenedies*/human
	Semantically related in English and Arabic	bee	*toroko*/sting

Table 4.9 The masked priming lists for passage B

List	Priming conditions	Targets	
		Prime	**Target/Meaning**
List 1	Semantically related via the Arabic root	travel	*sataca*/trip
	Semantically unrelated real word	internet	*pocoko*/suburb
List 2	Semantically related in English and Arabic	picnic	*sataca*/trip
	Non-word	hubana	*pocoko*/suburb
List 3	Semantically unrelated real word	minute	*sataca*/trip
	Semantically related via the Arabic root	suburb	*pocoko*/suburb
List 4	Non-word	palino	*sataca*/trip
	Semantically related in English and Arabic	town	*pocoko*/suburb

The number of distractor items in each list was 30; the distractor non-words were not encountered in any reading task (see Table 4.10). These were categorised into five groups: non-words primed by non-words, real words primed by semantically related real words, real words primed by semantically unrelated real words, real words primed by non-words, and non-words primed by real words (see Appendix 16a). Ten practice trials were introduced just before List 1 to make sure that participants became familiar with the task and that their reaction times were valid from the first experimental trial. For each passage, then, the SPT consisted of 32 items in each list, including the two non-words. The total number of items in the four lists was 138, including the ten practice trials. The number of distractors used in this type of experimental paradigm was large enough so that participants' attention was not drawn to the target items too easily. The order of items in the four lists was systematically randomised: List 2 was the reverse order of List 1; List 3 contained the second half of List 1 reversed, followed by the first half of List 1 reversed; List 4 was the reverse order of List 3. The items in the lists were also carefully organised so that the target non-words occurred at least ten items apart and there were no more than three non-words presented together or three real words shown together.

Table 4.10 The number of items in the SPT of each passage (in each list)

Primes	Targets	
	Real words	**Non-words**
Non-word	4	10
Semantically related real words	4	0
Semantically unrelated real words	4	8

All the SPT stimuli were controlled for length as far as possible, and presented individually in the middle of the screen using black 18-point Times New Roman font against a white background. The participants were instructed as follows: “Words will appear in separate trials in the middle of your screen. Each trial will include two sequential stimuli exposed for a fixed time (word then mask) and in the third exposure you should decide whether you know the meaning of the word as quickly and as accurately as possible. Press 'M' if you do (know the meaning), and press 'X' if not (do not know the meaning). You will see a '*' in the centre of the screen for 20 seconds before the first trial is presented. During this time, please place a finger of your left hand on the 'X' key and a finger of your right hand on the 'M' key, so that you are ready to start when the '*' disappears. You will have 4 practice trials to familiarise yourself with the procedure. Press the 'Space Bar' when you are ready to start” (see Appendix 16b). The participants were given a short break of a minute or two between the lists. To make this possible, a slide was inserted into the experiment with the text: “Please take a short break now. Press the space bar when you are ready to continue.” The same lists for both passages (A and B) with the same order were used one week later for the delayed post-test.

4.3.8 Digit Span Test (DST)

Forward and backward digit span is one of the most widely used neuropsychological tests of working memory (WM) (Richardson, 2007). It was originally a subtest of the widely used Wechsler Memory Scales (WMS) and the Wechsler-Adult Intelligence Scale (WAIS-III) (Wechsler, 1997). As reported by Wechsler (1997), the digit span test can be one of the best standardised measures of verbal working memory, particularly the backward digit span test, as the test-retest reliability is estimated as 0.83. As mentioned in Chapter 2 (Section 2.11.6), the Digit Span Test has been commonly used independently of general intelligence tests to examine participants’ working memory (WM) capacity. This test, therefore, seems to be consistent with the condition of the current study because it is needed for the comparison of group data to ascertain whether or not the two groups’ scores in all tasks were affected by differences in their WM capacity, as it is possible that L2 participants’ performance may be affected by their increased memory storage ability. This was proposed by Papagno and Vallar (1995), who suggested that L2 participants’ word learning is rooted in their higher working memory ability. Further research (Gupta, 2003) has shown a relationship between word learning and working memory performance. Moreover, individual differences in WM capacity may also reflect differences in the online task (SPT) performance, which has heavy processing requirements. Hence, it is important to ascertain whether the two groups’ performance in the online task was comparable with respect to their memory capabilities.

4.3.8.1 Administering and scoring the test

Following the procedure as described by the WAIS-III Administration and Scoring Manual for adults (Wechsler, 1997) and the WAIS-IV Adult Intelligence Scale (Wechsler, 2008), the Digit Span Test was conducted verbally in the current study. The advantage of using verbal digit-span tasks is to make sure that all L2 participants conduct the test in the same language. If the test is offered in a visual format, participants may perform the test in their strongest language, but it is unknown to the researcher in which language participants store and process information in WM (Olsthoorn, Andringa, & Hulstijn, 2014). Researchers of digit span have shown that bilinguals' familiarity with language plays a central role in their digit span performance, as they tend to achieve higher scores in their mother tongue (Chincotta & Hoosain, 1995; Thorn & Gathercole, 2001). In the current study, hence, the DST was conducted in the participants' first language (Arabic) as well as their second language (English). This way, information about participants' ability to process digits in both languages was obtained.

The digit sequences used in the participants' L2 were different from the digit sequences used in the L1 trial. The oral presentation of the numbers in English and Arabic was recorded to ensure that each participant underwent the same procedure. Two lists of numbers in forward and backward sequences were created. The lists contained strings starting with three digits, followed by increasingly longer strings of digits. For instance, the first sequence started with three digits 3-9-5, the next one was 7-3-8-4, followed by 5-1-7-3-9, etc. (see Appendix 15a).

The participants were asked to listen carefully to each digit presented via the recorder and to repeat the numbers that were read to them. This process continued even if the repeated string of digits was only partially correct; consecutive trials were presented until the participant was unable to repeat any of the digits in the sequence accurately. In the reverse trial of the Digit Span, the recorder presented a series of numbers to participants and they were asked to repeat the sequence in reverse order. For example, when the sequence 4-2-7 was presented, the respondent had to repeat these in the following order: 7-2-4. This sequence was also continued until the participant made an error in all three digits of the full sequence of numbers. The participants' responses were also recorded to help the researcher check them carefully. The Digit Span test was scored by recording the longest digit string sequence the participant was able to remember in each test. The scorer added up the total number of correct sequences, backwards and forwards.

Prior to the analyses of the data from the online test, the results of the DST were analysed to investigate whether there were differences in participants who took the ET with non-words

and those who took the ET without non-words. The individual trials were analysed using the total raw scores of correct sequences, backwards and forwards in both languages. The results revealed no outliers and no differences between L2 groups with respect to their working memory ability (see Appendix 15b).

4.3.9 Procedure

4.3.9.1 Procedure (Part 1)

The participants were not told what the main purpose of the study was, in order to avoid intentional new vocabulary learning. As in Study 1, they were told that the study focused on reading comprehension in L1 Arabic learners of English and they were needed for the study (see Appendix 17). The participants firstly completed a brief questionnaire with information about their educational background, age and interests, because the researcher needed to obtain information about intervening factors such as age and proficiency which may have affected L2 learners' performance (see Appendix 18). The decoding test (SWRT) was administered to ensure that the YARC passage level (Level 1) was the appropriate level for all participants. The researcher instructed the participants to read each word aloud, and, where the learner did not know the word, to try to sound out the word; if the learner still did not know the word, they then moved on to the next word. Learners read the words as accurately as possible. The researcher was also allowed to help participants to read the words or tell them whether each word was right or wrong.

The participants were divided into two groups: G1 and G2. The G1 group were given the YARC Passage Reading A containing two non-words. Immediately after reading Passage A, the G1 group took the YARC comprehension test on that passage, which comprised two multiple-choice reading comprehension questions. They were allowed to refer back to the passage to answer the questions and were required to complete the YARC test within ten minutes. As Keating's (2008) results confirmed the importance of the amount of time taken to complete the task, time-on-task was carefully controlled in the current study. Then, they carried out the post-reading activity, the Elaboration Task (ET). In the ET, the G1 group was divided into two groups: one group received the elaboration task, which included non-words, and the other group received the distractor version that did not contain non-words. Because each task consisted of 32 sentences, both groups were required to spend no more than 32 minutes in completing these activities. Hence, approximately one minute was given to complete each sentence. Then, they received the SPT to measure the recognition of the meaning of the new words from Passage A. After a week's delay, as in the studies of Hulstijn and Laufer (2001),

and Yaqubi et al. (2010), the G1 group took the SPT to assess the retention of word meaning of the two non-words from Passage A, and the VKS, completed in 12 minutes, to assess the retention of those words (meaning and use in a sentence) (see Figure 4.1).

The G2 group were asked to read the YARC Passage Reading B, containing two non-words. Immediately after reading Passage B, the G2 group were given the two multiple-choice reading comprehension questions on that passage. They were allowed to refer back to the passage to answer the questions and were required to complete the YARC test within ten minutes. Then, they took the ET, in which they were divided into two groups: one group received the ET which included target non-words, and the other group received the distractor version that did not include target non-words. Both two groups were required to complete the exercise within 32 minutes. Then, they received the SPT to measure the learning of the meaning of new words from Passage B. A week later, the G2 group took the SPT to assess word meaning retention of the two target non-words from Passage B, and the VKS completed within 12 minutes, to assess how those words were retained (meaning and use) after a week's delay (see Figure 4.1).

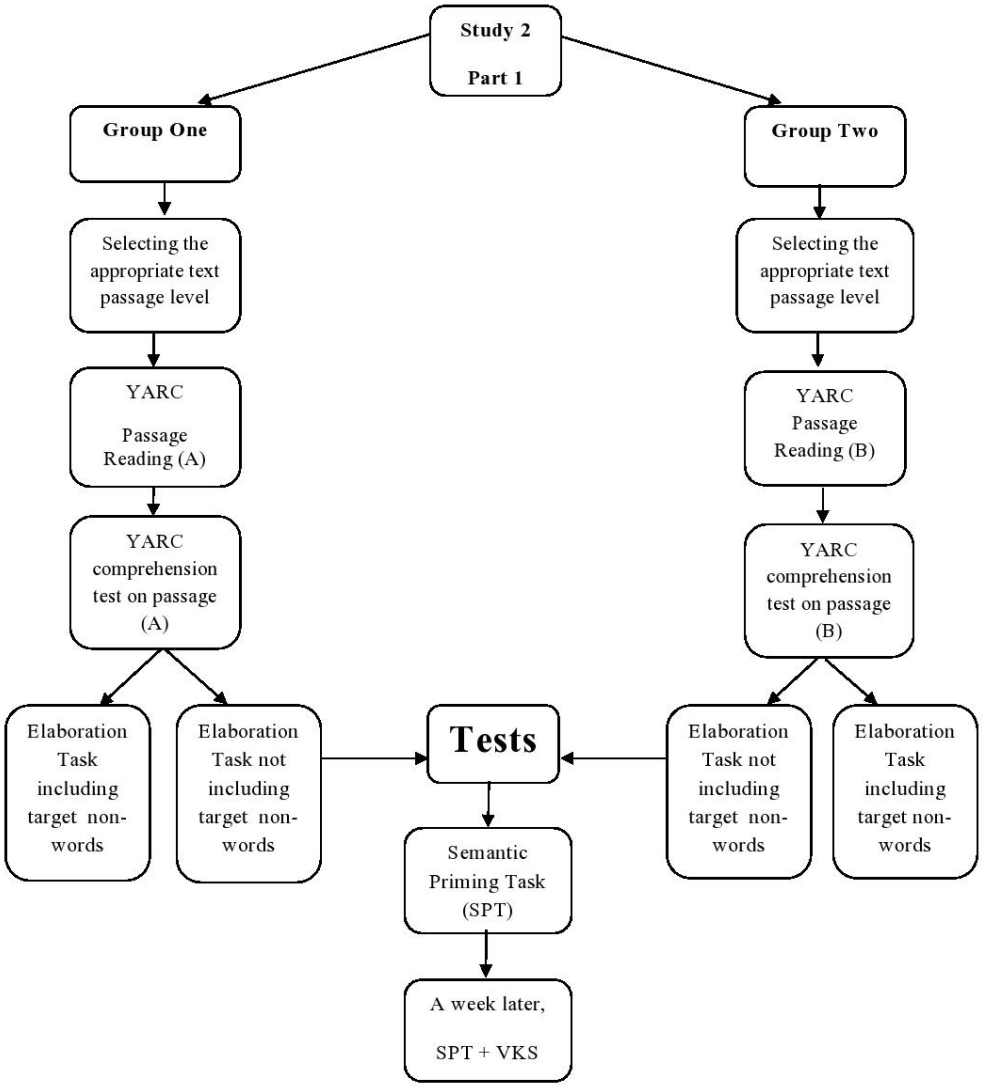


Figure 4.1 Procedure (Part 1)

4.3.9.2 Procedure (Part 2)

The design was counter-balanced so that G1 received the YARC Reading Passage B, while G2 received the YARC Reading Passage A. Part 1 and Part 2 of the study followed the same sequence of tasks for each group (see Figure 4.2). In the results section below, the scores of the two groups were combined and the analysis was based on the scores of the group who had carried out the elaboration task, including non-words, and the group who received the task that did not include non-words (see Section 4.4.1). At the end of the experiment, a Digit Span test was used in order to measure the participants' working memory capacity.

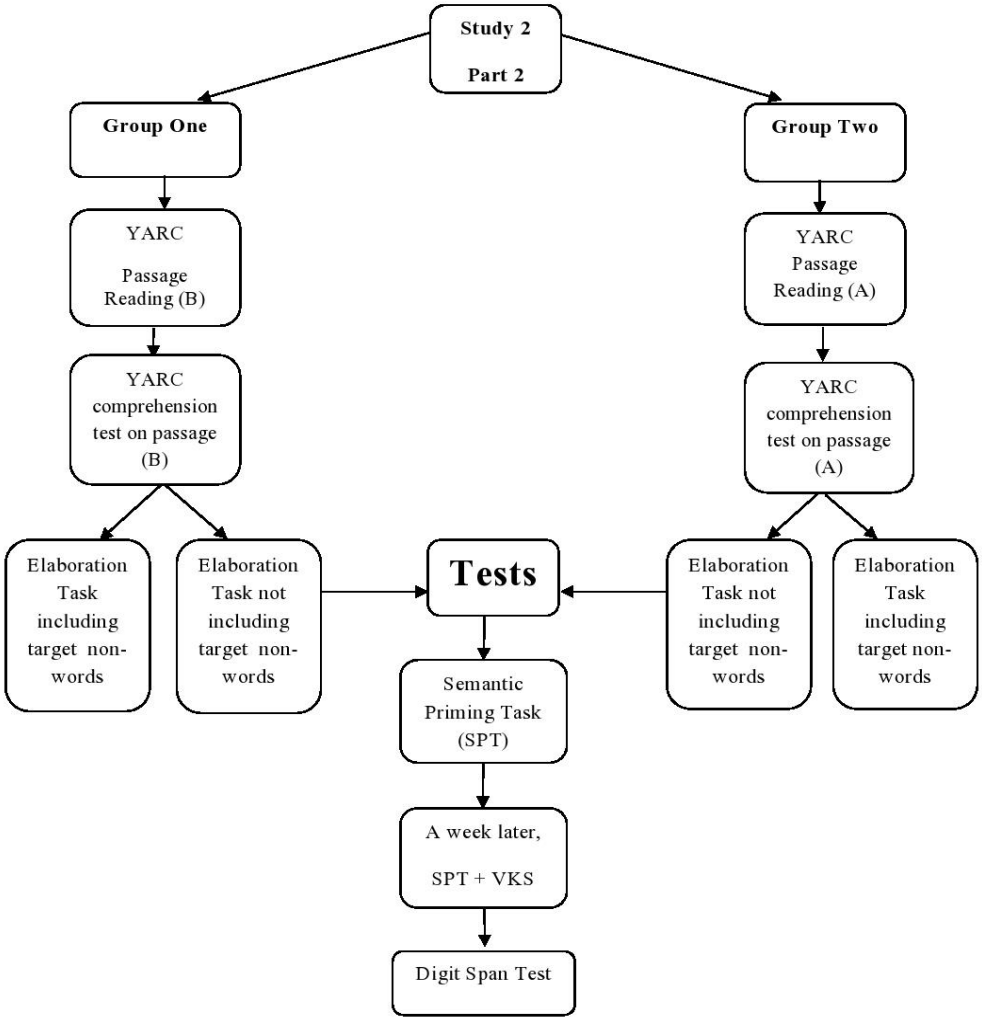


Figure 4.2 Procedure (Study 2, Part 2)

4.3.10 Ethics approval procedure

Prior to the data collection, ethical approval for this study was gained through the standard procedure required by the Department of English Language and Applied Linguistics at the University of Reading (see Appendix 19).

4.4 Analysis of Results

4.4.1 Introduction

This section presents the results of Study 2. The research question of this study was to investigate how an elaboration task (ET) can enhance L2 readers' retention of the meaning of new words encountered in a reading task. As explained in Section 4.3.5, the ET was developed as a post-reading activity to help participants reflect on the meaning of the target non-words presented in the stories. This study is based on Laufer and Hulstijn's (2001) Involvement Load Hypothesis (ILH), by examining to what extent a high involvement component (depth of processing levels) can relate to the retention of the meaning of new words.

Non-words were used as the target items for measuring participants' ability to learn and retain the meaning of newly learned items. Retention of meaning was tested with a Semantic Priming Task (SPT). The Vocabulary Knowledge Scale (VKS) was also used to test knowledge of meaning and use (see Instruments Section 4.3.2). A Kolmogorov-Smirnov test for normality was used to test all measures used in the study. Non-parametric tests were used to investigate whether the group differences were significant, because not all scores of these tests for L1 Arabic learners of English were normally distributed (see Appendix 20a). For each of the Mann Whitney U-tests and the Wilcoxon T tests reported in this section, the value of N = 40. The analysis was based on the performance of the group who had carried out the elaboration task with target non-words (ET- A1 or B1) and their performance on the task that did not contain target non-words (ET- A2 or B2) (see Section 4.3.5). Effect sizes were calculated to see the strength of significant differences achieved. According to Cohen's (1988) estimation of the effect size, d = 0.2 is a small, d = 0.5 is a medium, and d = 0.8 is a large effect size. The results of all tests: the York Assessment Reading Comprehension Secondary Test (YARC), ET, VKS, and SPT (IPT and DPT), are presented below, followed by the analysis of the tests and the correlations between them.

4.4.2 Comprehension scores

The reading comprehension performance was measured via the YARC Secondary Test, which included the Single Word Reading Test (SWRT) and multiple-choice reading comprehension questions (RCQ). The RCQ were designed to be relevant to the target non-words, on passages A and B. The mean accuracy scores of RCQ and SWRT are presented in Table 4.11.

Table 4.11 Mean accuracy scores and standard deviations on the YARC Secondary Test

N = 40	**RCQ Passages A and B Max = 4**	**SWRT Max= 70**
Mean	3.03	47.98
SD	0.92	2.50

4.4.3 Elaboration Task Scores

As mentioned in Section 4.3.5, there were four ET tasks: an ET which contained four or eight tokens of each of the two non-words from Passage A (A1) and an ET-distractor task with real words only (A2). In addition, there was an ET which contained four or eight tokens of each of the two target non-words from Passage B (B1), and an ET-distractor task with

real words only (B2). All 40 participants took the ETs, but the order of presentation was reversed; this means that the group who had received A1 also took B2 and the group who had taken A2 received B1, in order to avoid the repetition effects. Each of the ETs that included non-words (A1 and B1) was split into two main categories: 16 sentences containing non-words and 16 containing real words. The ET-distractor samples (A2 and B2) did not include sentences with non-words.

In order to be able to combine A1 with B1, the mean accuracy scores of the non-word sentences, as well as real-word sentences, were compared between A1 and B1. No significant differences were found in the Mann Whitney U-test between the scores. Another between-group comparison of real-word sentences was made between A2-B2, and the Mann Whitney U-test showed no significant differences between the scores (see Appendix 20b). Then, A1 was combined with B1 scores and A2 with B2 scores. Participants from the A2-B2 groups had not had any training in the use of the non-words through the ET.

Because the participants needed to judge how appropriate these sentences were on a six-point scale, the maximum value for the mean accuracy scores and SD reported in this section is 6. The mean accuracy scores of the sentences containing target non-words and real-word sentences in A1-B1 are presented in Table 4.12. A highly significant difference in accuracy was found ($Z = -5.515$, $p < 0.001$, $d = 3.56$), which shows that, as expected, the group performed better on the sentences with real words than on sentences with non-words.

Table 4.12 Mean accuracy scores and standard deviations on the ET (A1-B1) target non-words and real words

(N = 40)	**ET - target non-word sentences (A1-B1)**	**ET – real-word sentences (A1-B1)**	**Sig/Diff**
Mean	4.84 (80.60%)	5.44 (90.73%)	0.001
SD	0.79	0.47	

The A1 and B1 categories were also divided into three types of sentences: 8 grammatically and semantically appropriate real- word sentences (category 1), 8 grammatically and semantically appropriate target non-word sentences (category 2), 4 semantically anomalous real word sentences (category 3), 4 semantically anomalous target non-word sentences (category 4), 4 grammatically incorrect real-word sentences (category 5), and 4 grammatically incorrect target non-word sentences (category 6). The sentences in A2 and B2 were also divided into three types: 16 grammatically and semantically appropriate real-

word sentences, 8 semantically anomalous real-word sentences, and 8 grammatically incorrect real-word sentences.

The mean accuracy scores of A1-B1 and A2-B2 categories are presented in Table 4.13. In a comparison of real-word sentences between A1-B1 and A2-B2, highly significant differences in accuracy were found in category (1) (Z = -3.922, p < 0.001, d = 1.58), in category (3) (Z = -4.054, p < 0.001, d = 1.67), as well as in category (5) (Z = -3.661, p < 0.001, d = 1.42). This indicates that the participants obtained higher scores in all real-word categories in A1-B1 than in A2-B2 (see Table 4.13). This may have occurred because A1-B1 involved target non-word sentences, which may have helped the participants to concentrate more on the real-word sentences and then to perform better in ETs A1-B1 than A2-B2.

In a comparison between the accuracy scores for different types of sentences with real words and sentences with non-words of A1-B1, there were also highly significant differences, between categories 1 and 2 (Z = -2.961, p< 0.003, d = 1.06), between category 3 and 4 (Z = -2.620, p < 0.009, d = 0.91), and between category 5 and 6 (Z = -2.588, p < 0.01, d = 0.90), which indicates that participants obtained higher scores in all three types of real-word sentences than sentences containing non-words. This means that they were better at answering first the semantically anomalous real-word sentences, then semantically anomalous non-word sentences, then grammatically incorrect real-word sentences, and last, grammatically incorrect non-word sentences (see Table 4.13).

Table 4.13 Mean accuracy scores and standard deviations for real words and target non-words from A1 and B1, and for real words from A2 and B2

N = 40	ET (A1-B1)		ET (A2-B2)		Sig/ Diff
	Mean	SD	Mean	SD	
Real words					
Grammatically and semantically appropriate real words	5.5 (91.67%)	0.61	4.47 (74.44%)	0.97	0.001
Semantically anomalous real words	5.90 (98.25%)	0.36	5.66 (94.33%)	0.65	0.001
Grammatically incorrect real words	4.88 (81.38%)	1.32	4.14 (69.06%)	1.11	0.001
Target non-words					
Grammatically and semantically appropriate -non-words	4.96 (82.60%)	1.14	_____	_____	_____
Semantically anomalous non-words	5.31 (88.46%)	1.19	___	_____	_____
Grammatically incorrect non-words	4.13 (68.75%)	1.43	_____	_____	_____

Overall, it seems that participants performed better on the sentences with real words than on sentences with target non-words. The results also showed that in all three types of real-word sentences of A1-B1, they obtained higher scores than in the A2-B2. In a comparison between real words and target non-word sentences of A1-B1, the results showed, as expected, that they were better at answering all three types of real-word sentences than target non-word sentences.

4.4.4 Vocabulary knowledge Scale (VKS) scores

As the main interest of the study was the retention of target non-words, the VKS results were subsequently split into 8 real words and 4 target non-words. The difference in accuracy between real words and target non-word categories was highly significant (Z = -4.761, $p < 0.001$, d = 2.29), as the mean accuracy percentage of the real words was higher than it was for the target words (see Table 4.14).

Table 4.14 Mean accuracy percentages and standard deviations on the VKS target non-words and real words

Non-words (Max = 8)		Real words (Max = 16)		Sig/Diff
Mean	**SD**	**Mean**	**SD**	0.001
69.06%	28.16	99.06%	4.37	

The VKS results were also divided into two main categories: one relating to the group who had received the ET, including target non-words from the two passages (A1 or B1) prior to taking the VKS, and the other for the same group when they had received the version with distractors for either of the two passages (A2 or B2). Firstly, a comparison in VKS scores between A1 and B1 was made: no significant differences were found between VKS scores whether participants had taken A1 or B1. In addition, no significant differences were found between A2 and B2 VKS scores. Then, VKS-A1 was combined with VKS-B1 scores and VKS-A2 with VKS-B2 scores.

Table 4.15 presents the mean accuracy scores of the VKS scores for the participants who had taken A1 or B1 and those who had been given A2 or B2. It was expected (Prediction 1 of RQ1) that the VKS scores would be higher for the group which had carried out the elaboration task containing these words than for a control group. There was a highly significant difference in accuracy (Z = -4.471, $p < 0.001$, d = 2.00), which shows that participants who had taken the ET- A1 or B1 achieved higher scores in the VKS than those who had completed the ET-distractor version A2 or B2.

Table 4.15 Mean accuracy scores and standard deviations on the VKS

N = 40 Max = 12				
VKS/ET (A1 or B1)		VKS/ET- (A2 or B2)		Sig/Diff
Mean	**SD**	**Mean**	**SD**	
11.60 96.67%	1.13	9.78 81.5%	1.76	0.001

The mean accuracy scores of the VKS real words and target non-word categories for the participants who had also taken A1-B1 or A2-B2 are presented in Table 4.13. There were highly significant differences in accuracy in the non-word category (Z = -4.565, $p < 0.001$,

d = 2.07), which shows, as expected, that they achieved higher scores when they had taken the ET with non-words (A1 or B1) than when they had taken the distractor version (A2 or B2). By contrast, for the real words, there was no significant difference in accuracy between their performances (see Table 4.16).

Because the VKS provides information on meaning as well as use of the words in the study, the VKS scores on target non-words were also categorised into meaning scores and usage scores (see Table 4.16). For providing the meaning of the non-words, there was a highly significant difference in accuracy (Z = -3.938, p < 0.001, d = 1.59). Another highly significant difference in accuracy was found in relation to the use of the non-words in a sentence (Z = -4.636, p < 0.001, d = 2.16). This indicates that participants showed better performance after receiving the ET, including target non-words (A1 or B1).

Table 4.16 Mean accuracy scores and standard deviations on the VKS categories

	(N = 40)				
	VKS/ET (A1 or B1)		**VKS/ET (A2 or B2)**		**Sig/Diff**
	Mean	**SD**	**Mean**	**SD**	
Real words (Max = 8)	7.95	0.32	7.90	0.63	NS
Non-words (Max = 4)	3.65	1.10	1.88	1.59	0.001
Meaning of the non-words (Max = 2)	1.83	0.55	1.13	0.88	0.001
Using the non-words in sentences (Max = 2)	1.83	0.55	0.75	0.87	0.001

Overall, it seems from the general accuracy scores that participants were better at retaining real words than target non-words. The general VKS results indicated, as expected, that participants who took the ET with target non-words from the two stories achieved higher scores in the VKS than those who filled in the ET-distractor version. In terms of the mean accuracy scores of target non-words, and, particularly, in providing the meaning and how to write the non-words in a sentence, participants who took the ET with target non-words from both stories outperformed those who had been given the ET-distractor version.

4.4.5 Semantic Priming Task (SPT) accuracy scores and RTs

All participants received the SPT (IPT and DPT) in order to test their retention of the meaning of new words. As mentioned in Section 4.3.7, the SPT in both passages consisted of

16 non-words and 96 real words presented as distractors. The differences in accuracy between real words and non-words were highly significant in the SPT (IPT) (Z = -5.139, p < 0.001, d = 2.79), as well as the SPT (DPT) (Z = -4.105, p < 0.001, d = 1.71), as the mean accuracy percentage for real words was higher than for the non-words (see Table 4.17). It is also worth mentioning that the mean accuracy scores of both real words and target non-words in the DPT were higher than in the IPT. The differences in accuracy between IPT and DPT were highly significant for real words (Z = -2.998, p = 0.003, d = 1.08), as well as for target non-words (Z = -3.339, p = 0.001, d = 1.24). This indicates that participants learned the target non-words as well as being able to retain these newly learned words in the DPT.

Table 4.17 Mean accuracy scores and standard deviations on the SPT (IPT and DPT) target non-words and real words

SPT N = 40	Non-words (Max = 16)		Real words (Max = 96)		Sig/Diff
	Mean	SD	Mean	SD	
IPT	70.00%	25.72	95.26%	4.60	0.001
DPT	80.94%	23.64	96.82%	4.79	0.001

The participants' SPT (IPT and DPT) mean scores were also divided according to the ET types: when the group had received the ET with target non-words from the two passages (A1 or B1), and when they had been given the ET-distractor versions for either of the two passages (A2 or B2). Firstly, a comparison in SPT (IPT and DPT) scores between A1 and B1 was made: no significant differences were found between SPT (IPT and DPT) scores, whether participants had taken A1 or B1. Another comparison was made in SPT (IPT and DPT) scores between A2 and B2, and again no significant differences were obtained. Then, SPT-A1 was combined with SPT-B1 scores and SPT-A2 with SPT-B2 for both IPT and DPT scores. It was expected (Prediction 1 of RQ1) that the Semantic Priming Task scores would demonstrate better retention of the meaning of target non-words in the group which had carried out the elaboration task with the target non-words. The results showed that there were no significant differences in accuracy on SPT for either the IPT or the DPT, which indicates that participants showed similar performance in the SPT both (IPT and DPT) whether or not they had taken the ET with non-words.

In terms of the speed of processing RTs, all error trials were excluded from the reaction time data. Table 4.18 shows the RT means of target non-words and real words (distractors) on

the SPT (IPT and DPT). As expected, the RTs of the real words were shorter than the RTs for target non-words on the SPT both (IPT and DPT) (see Table 4.18). The differences in RTs between real words and target non-words were highly significant for the SPT (IPT) (Z = -4.633, p = 0.001, d = 2.15), as well as for the SPT (DPT) (Z = -4.516, p = 0.001, d = 2.04). It is also worth mentioning that the response time of both real words and target non-words in the DPT was shorter than in the IPT. The RT differences between IPT and DPT were highly significant for real words (Z = -2.621, p = 0.009, d = 0.91), as well as for non-words (Z = -2.791, p = 0.005, d = 0.98).

Table 4.18 Means and standard deviations of the RTs for target non-words and real words on the SPT (IPT and DPT)

SPT N = 40	Target non-words		Real words		Sig/Diff
	Mean	SD	Mean	SD	
IPT	906.05	233.19	773.81	133.82	0.001
DPT	842.03	179.34	744.00	109.27	0.001

The participants' SPT (IPT and DPT) RT means were also divided according to the ET types: when the group had taken the ET with target non-words from the two passages (A1 or B1), and when they had been given the ET-distractor versions for either of the two passages (A2 or B2). Firstly, a comparison in RTs from the SPTs was made for participants who had taken either the A1 or the B1 ET. This did not result in significant differences in response time between both SPTs (IPT and DPT). The conclusion was therefore drawn that ET A1 and B1 affected the RTs for the SPTs in the same way. Another comparison was made between RTs on the SPT (IPT and DPT) for participants who had taken either A2 or B2. Again, no significant differences were found. It is therefore concluded that the ET distractor tasks A2 and B2 affected the RTs in the SPT in similar ways. Then, the RTs from the SPT were compared for the groups who had taken the ET with target non-words (either A1 or B1) and the groups who had taken the distractor ET with real words (either A2 or B2). Participants from the latter groups had not had any training in the use of the non-words through the ET.

It was predicted (Prediction 1 of RQ1) that the reaction time (RT) for target non-words would be faster for the group which had carried out the elaboration task with the target non-words. Based on the RT means of SPTs (IPT and DPT), the results showed that there were no significant RT differences for these two groups. Whether they had taken the ET with target non-words or not did not enhance the RTs on the SPT (see Appendix 20f).

Overall, it seems from the general percentage accuracy scores, that participants were better at retaining real words than target non-words, and the response time for the real words was also shorter than for target non-words in both the IPT and the DPT of the SPT. It was also found that for both real words and target non-words, the participants' accuracy scores on the DPT were higher than on the IPT, and also their processing speed in the DPT was shorter than in the IPT. No significant differences were found between groups taking ETs with or without target non-words, as shown by the accuracy scores and RTs for SPTs (IPT and DPT).

4.4.6 The contribution of depth of processing to improving L2 readers' retention of the meaning of new words

The first research question of this study was to investigate whether a task with a high involvement component, the ET, could contribute to the retention of the meaning of new words encountered in a reading task. The independent variables were the mean target non-word scores for the ET in the two passages (A1-B1). The dependent variables were the scores on the SPT (IPT), the SPT (DPT) and the VKS.

From the results presented in Section 4.4.4, it was clear that the ET with the target non-words had enhanced the retention of the non-words in the VKS scores. However, from the mean accuracy scores of SPT (IPT and DPT) shown in Section 4.4.5, it seems that the participants showed similar performance in the SPT (IPT and DPT), whether or not they had taken the ET with target non-words. As can be noticed, the results obtained from the VKS were different from those found on the SPTs, which will be further clarified in the discussion chapter.

Spearman Correlations were computed to examine how the ET target non-word scores related to the non-word scores of VKS, and non-word scores of SPT (IPT and DPT). There was a highly significant correlation between ET non-word scores and SPT (IPT) target non-word scores ($r = 0.540$, $p < 0.01$) as well as between ET non-word scores and SPT (DPT) target non-word scores ($r = 0.631$, $p < 0.01$) (see Table 4.19).

Table 4.19 Spearman Correlations between target non-word scores of ET (A1 and B1), target non-word scores of VKS, SPT (IPT and DPT) for both passages

N = 40	ET target non-word scores	VKS target non-word scores	SPT (IPT) target non-word scores	SPT (DPT) target non-word scores
ET target non-word scores	___			
VKS target non-word scores	0.227	___		
SPT (IPT) target non-word scores	0.540**	0.306	___	
SPT (DPT) target non-word scores	0.631**	0.275	0.592**	___

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

Three regression analyses were computed to investigate the association of administering the ET with learning and retention of the meaning of new words encountered in a reading task. First of all, a regression analysis was computed with the target non-word scores of VKS as the dependent variable, then with target non-word scores of SPT (IPT), and, lastly, with the target non-word scores of SPT (DPT) as the dependent variables. The independent variable was the target non-word scores of ET (A1-B1). The regression model suggested that the learners' ET non-word scores did significantly associate with the target non-word scores in the VKS (B = 0.311, p = 0.05). However, the explanatory power of the model was limited, as suggested by an adjusted R square value of 0.096. This means that 9.6% of the variance in VKS non-word scores was accounted for by the ET non-word scores (see Appendix 20c).

It was also suggested that the learners' ET non-word scores significantly related to the target non-word scores in the SPT (IPT) (B = 0.544, p = 0.001). The model had a stronger explanatory power, as suggested by an adjusted R square value of 0.296. This means that 29.6% of the variance in SPT (IPT) non-word scores was accounted for by the ET non-word scores (see Appendix 20d). Lastly, the regression model also suggested that the learners' ET non-word scores significantly associated with the target non-word scores in the SPT (DPT) (B = 0.602, p = 0.001). The model had an explanatory power, as suggested by an adjusted R square value of 0.363. This means that 36.3% of the variance in SPT (DPT) non-word scores was accounted for by the ET non-word scores (see Appendix 20e). This clearly contrasts with the findings mentioned previously (in Section 4.4.5) which show the absence of

significant differences on the SPTs between groups taking ETs with or without target non-words. Nevertheless, the results suggested by the regression model seem to provide strong evidence for the association of the ET on the retention of the target non-words. Therefore, Hypothesis 1 was supported by the findings, since it seems from the results shown above, that the ET with the target non-words had a positive relationship with the participants' target non-word scores on the VKS and the SPT (IPT and DPT).

4.4.7 Priming conditions

The second research question of this study was to explore to what extent cross-language semantic priming enhances word recognition in L2 learners of English. Four types of primes were used with each target non-word: Arabic root prime (1), Arabic and English prime (2), unrelated word prime (3), and unrelated non-word prime (4) (see Section 4.3.7). It was predicted (Prediction 1 of RQ2) that L2 learners would be faster at recognising L2 words from L2 semantically related prime words. With the two conditions of the target non-words, it was possible to identify the baseline measure for semantic relatedness, i.e., to determine whether participants had started to build lemmas with L2 word meanings or whether they continued to map the L1 lemma content onto the target L2 form. If participants had acquired the target meaning in English, it was predicted that accuracy scores in the SPT (IPT and DPT) would be higher for condition 2 than for condition 1, and if they were still relying solely on the Arabic lemma, the highest scores would be obtained by condition 1, followed by condition 2.

The mean accuracy scores of the four conditions of non-words in SPT (IPT) in the two passages are presented in Table 4.20. There was a significant difference in accuracy between 1 and 2 (Z = -2.129, p < 0.03, d = 0.72), as the mean accuracy scores in condition 2 were significantly higher than they were in condition 1. This shows that participants' responses were facilitated by the Arabic and English prime and not solely by the Arabic prime. Accuracy rates were significantly higher in condition 3 than 1 (Z = -2.619, p < 0.009, d = 0.91); another significant difference in accuracy rates was that they were higher in condition 4 than in 3 (Z = 2.200, p = 0.02, d = 0.74). The overall results, after comparing the four conditions of the target non-words for the two stories, were as follows: 1 < 2 = 3 < 4 (see Table 4.20).

Table 4.20 Mean accuracy scores and standard deviations for different priming conditions in the SPT (IPT)

N = 40 Max = 4	Passages A&B	
	Mean	SD
Arabic root prime (1)	2.48 62.00%	1.20
Arabic & English prime (2)	2.75 68.75%	1.19
Unrelated word prime (3)	2.85 71.25%	1.19
Unrelated non-word prime (4)	3.13 78.25%	0.99

The mean accuracy scores of the four conditions of target non-words in the SPT (DPT) in both passages are presented in Table 4.21. Accuracy scores were just significantly higher in condition 1 than condition 3 (Z = -1.999, p < 0.04, d = 0.67), but the difference between conditions 1 and 2 failed to reach significance. This indicates that, in terms of accuracy, the ability of the Arabic and English primes to facilitate recognition of the meaning of target words, observed at IPT, had disappeared. The overall results after comparing the four conditions of the target non-words in the DPT were as follows: 1 = 2 = 3 < 4 (see Table 4.21).

Table 4.21 Mean accuracy scores and standard deviations for different priming conditions in the SPT (DPT)

N = 40	Passages A & B	
	Mean	SD
Arabic root prime (1)	3.36 84.00%	0.90
Arabic & English prime (2)	3.27 81.75%	1.07
Unrelated word prime (3)	3.13 78.25%	1.19
Unrelated non-word prime (4)	3.40 85.00%	0.93

With respect to the RTs, table 4.22 presents the RT means of the four priming types used with each target non-word in SPT (IPT and DPT). As mentioned in Section 4.3.7, if participants had started to build lemmas with L2 word meanings, it was predicted that the RT on the SPT (IPT and DPT) would be 2 < 1 and if they were still relying on the Arabic representation of the target, then the following order was predicted: 1 < 2. This predication was not supported by the RT data (IPT), in contrast to the accuracy data. The only significant difference was obtained between the RT means of Conditions 3 and 4 (Z = -2.226, p = 0.026,

d = 0.75). The overall rank order of RTs, according to the four priming conditions in the SPT (IPT), was as follows: 1 = 2 = 3 > 4 (see Table 4.22).

Table 4.22 Means, standard deviations and percentage accuracy scores of the RTs for different priming conditions in the SPT (IPT)

N = 40	Passages A & B	
	Mean	**SD**
Arabic root prime (1)	909.09 62.00%	242.53
Arabic & English prime (2)	907.17 68.75%	392.64
Unrelated word prime (3)	918.42 71.25%	234.27
Unrelated non-word prime (4)	859.41 78.25%	190.15

The RT means of the four conditions of target non-words in the SPT (DPT) in both passages are presented in Table 4.23. The RT difference between 2 and 1 was highly significant (Z = -3.314, p = 0.001, d = 1.23), thus supporting the prediction that participants more likely to be relying on L2 integration of the lemma than L1 mediation in the delayed post-test. The RT mean of 2 was also shorter than the RT mean for 3 (Z = -2.444, p = 0.01, d = 0.84). The overall rank order of the RTs after comparing the four conditions of the non-words RT in the DPT was as follows: 1 > 2 < 3 = 4 (see Table 4.23).

Table 4.23 Means, standard deviations and percentage accuracy scores of the RTs for different priming conditions in the SPT (DPT)

N = 40	Passages A & B	
	Mean	**SD**
Arabic root prime (1)	880.98 84.00%	208.92
Arabic & English prime (2)	798.93 81.75%	210.64
Unrelated word prime (3)	850.52 78.25%	197.71
Unrelated non-word prime (4)	817.88 85.00%	207.34

Taking these findings together, it appears that in terms of accuracy, there was evidence at IPT that participants were not solely reliant on the Arabic prime, but this effect disappeared at DPT, whereas with respect to the RTs, no effect was found at IPT, but there was evidence that they had started to build lemmas with L2 word semantics at DPT.

It was also predicted (Prediction 2 of RQ2) that the group which had carried out an elaboration task with target non-words would succeed in obtaining a priming effect from L2 semantically related prime words. Tables 4.24 and 4.25 represent the accuracy scores of the four conditions of non-words according to the ET types: when the group had taken the ET with target non-words from the two passages (A1 or B1), and when they had received the ET-distractor versions for either of the two passages (A2 or B2). Accuracy scores were marginally significantly higher in condition 2 than condition 1 for the group who had taken the ET with non-words in the SPT (IPT) (Z = -1.897, p = 0.05, d = 0.63). This indicates, as predicted, that participants' responses were facilitated by the Arabic and English prime and not solely by the Arabic prime. Accuracy scores were significantly higher in condition 3 than condition 1 for the same group in the SPT (IPT) (Z = -2.138, p = 0.03, d = 0.76). The overall results, after comparing the four conditions of the target non-words for the ET non-word group in the IPT, were as follows: 1 < 2 = 3 = 4.

No differences were found between any of the four priming conditions of the non-words in the IPT for the group who had taken the ET-distractor version (see Table 4.24). These results indicate that in the IPT, the ET for non-words facilitated acquisition of those target words, whereas the distractor ET had no effect on performance.

Table 4.24 Mean accuracy scores and standard deviations for the SPT (IPT) four priming conditions of the target non-words by elaboration task type

(N = 40)	**SPT/ET A1 or B1**		**SPT/ET A2 or B2**	
	Mean	**SD**	**Mean**	**SD**
Arabic root prime (1)	1.31 65.50%	0.73	1.20 60.00%	0.78
Arabic & English prime (2)	1.48 74.00%	0.72	1.28 64.00%	0.75
Unrelated word prime (3)	1.50 75.00%	0.72	1.35 67.50%	0.70
Unrelated non-word prime (4) (Max = 2)	1.58 79.00%	0.71	1.55 77.50%	0.68

As presented in Table 4.25, the results revealed that accuracy scores in condition 1 were higher than those for condition 2 for the group who had taken the ET with target non-words in the

DPT. An analysis with the Wilcoxon's Signed Ranks test showed that these differences were significant (Z = -2.309, p = 0.02, d = 0.78). Accuracy scores were significantly higher in condition 1 than condition 3 (Z = -1.999, p < 0.04, d = 0.67) for the same group. This indicates that the group who had taken the ET with non-words were relying on the Arabic representation of the target meaning of the target non-words in the DPT, in contrast to their performance patterns at IPT, as shown in Table 4.25. Accuracy scores were just significantly higher in condition 4 than condition 3 (Z = -2.111, p = 0.03, d = 0.71) for the same group. The overall results, after comparing the four conditions of the target non-words for the ET non- word group in the DPT were as follows: 1 > 2 = 3 < 4 (see Table 4.25).

No differences were found between conditions 1 and 2 in the DPT for the group who had taken the ET- distractor version, which confirmed that the distractor ET had no effect on performance scores. A significant difference in accuracy was found (Z = -2.121, p = 0.03, d = 0.71) between 3 and 4 in the DPT for the group who had taken the ET-distractor version (see Table 4.25). The overall results, after comparing the four conditions of the target non-words for the ET-distractor version group in the DPT, were as follows: 1 = 2 = 3 < 4.

Table 4.25 Mean accuracy scores and standard deviations for the SPT (DPT) four priming conditions of the target non-words by elaboration task

(N = 40)	**SPT/ET A1 or B1**		**SPT/ET A2 or B2**	
	Mean	**SD**	**Mean**	**SD**
Arabic root prime (1)	1.79 89.50%	0.52	1.58 79.00%	0.71
Arabic & English prime (2)	1.60 80%	0.63	1.55 77.50%	0.75
Unrelated word prime (3)	1.60 80%	0.71	1.48 74%	0.78
Unrelated non-word prime (4) (Max = 2)	1.78 89.00%	0.53	1.63 81.50%	0.67

Tables 4.26 and 4.27 represent the RT means of the four conditions of target non-words according to the ET types: when the group had taken the ET with non-words from the two passages (A1 or B1), and when they had been given the ET-distractor versions for either of the two passages (A2 or B2). The difference in RTs between 2 and 1 was highly significant for the group who had taken the ET with non-words in the SPT (IPT) (Z = -2.627, p = 0.009, d = 0.91), which showed that the RT mean of 2 was shorter than the RT mean for 1. This indicates that the group who had carried out the ET with target non-words were relying not only on the Arabic

semantic system but the Arabic and English primes facilitated recognition of the meaning of target words encountered at the time of the IPT. The overall results after comparing the four conditions of the target non-words RT for the ET non-word group in the IPT were as follows: 1 > 2 = 3 = 4. No differences were found in RT means between all four priming conditions of the non-words in the IPT for the group who had taken the ET-distractor version (see Table 4.26).

Table 4.26 Means, standard deviations and percentage accuracy scores of the RT for the SPT (IPT) four priming conditions of the target non-words by elaboration task

(N = 40)	SPT/ET A1 or B1		SPT/ET A2 or B2	
	Mean	**SD**	**Mean**	**SD**
Arabic root prime (1)	937.30 65.00%	262.53	884.56 60.00%	300.44
Arabic & English prime (2)	879.70 74.00%	583.84	917.13 64.00%	314.86
Unrelated word prime (3)	890.96 75.00%	286.92	948.36 67.50%	290.26
Unrelated non-word prime (4) (Max = 2)	817.29 79.00%	226.27	883.90 77.50%	226.56

A significant difference was found between 2 and 1 for the group who had taken the ET with target non-words in the DPT (Z = -3.342, p = 0.001, d = 1.25), which showed that the RT mean of 2 was shorter than the RT mean for 1. The RT mean of 2 was also shorter than the RT mean for 3 (Z = -2.899, p = 0.004, d = 1.03). This also indicates that the group who had carried out the ET with target non-words had started to build lemmas with L2 word meanings at the time of the DPT. There was also no difference in RT mean between 3 and 4 for the same group. The overall rank order of RTs according to the four priming conditions was as follows: 1 > 2 < 3 = 4. The results showed that no differences were found in RT means between all four priming conditions of the target non-words for the group who had taken the ET-distractor version (see Table 4.27).

Table 4.27 Means, standard deviations and percentage accuracy scores of the RT for the SPT (DPT) four priming conditions of the target non-words by elaboration task

(N = 40)	SPT/ET A1 or B1		SPT/ET A2 or B2	
	Mean	SD	Mean	SD
Arabic root prime (1)	916.55 89.50%	255.70	808.26 79.00%	200.14
Arabic & English prime (2)	794.46 80.00%	237.23	796.79 77.50%	163.39
Unrelated word prime (3)	855.32 80.00%	214.32	823.36 74.00%	261.22
Unrelated non-word prime (4) (Max = 2)	809.87 89.00%	208.18	813.63 81.50%	233.78

Overall, for the four conditions of the target non-words, the accuracy scores indicated that participants had started to build lemmas with L2 word semantics in the SPT (IPT), but this effect disappeared in the DPT. The DPT RT means further indicated that participants were more likely to be relying on L2 integration of the lemma than L1 mediation, whereas there was no evidence of this finding in the IPT. An examination of the effect of the different priming categories in the elaboration task revealed that in the IPT, the group who had taken the ET with target non-words produced more accurate responses when an Arabic and English prime was used, indicating that they had started to build lemmas with L2 word meanings, but this effect was reversed at DPT. By contrast, the distractor version group did not result in significant differences in any of the four priming conditions of the target non-words in the IPT and the DPT. Therefore, in terms of accuracy, the overall effect of facilitating the Arabic and English prime was solely attributable to the effectiveness of the ET involving the target non-words at IPT.

The RT results showed that the group who had taken the ET with non-words were relying not only on the Arabic semantic system but also the ability of the Arabic and English primes to facilitate recognition of the meaning of target words observed at both IPT and DPT. The distractor version group, on the other hand, did not yield significant differences in any of the four conditions of the target non-words in the IPT or the DPT. Overall, all the RT results reported here indicate that participants who received the non-word ET had started building lemmas with L2 word semantics in both the IPT and the DPT, and that the distractor ET had no effect on processing times in the IPT or DPT.

4.4.8 The contribution of frequency to L2 vocabulary acquisition

The third research question of the current study aimed to investigate whether frequency of occurrence of the target words or frequency of the original words which had been replaced by target non-words explained the learning and retention of these words. The four target non-words were controlled by their occurrence in the texts and categorised into two sets: two non-words which occurred eight times (*cenedies* in Passage A and *pocoko* in Passage B), and two non-words which occurred four times (*toroko* in Passage A and *sataca* in Passage B). It was expected (Prediction 1 of RQ3) that target non-words which occurred more frequently in the texts would be better learned and retained, as evidenced by higher accuracy scores on the post-tests. Tables 4.28 reveal the mean accuracy scores of the non-words that occurred eight times and four times on the VKS. Although the scores appeared to be slightly higher for the words which occurred eight times, there was no difference between the scores for both groups of words on the VKS. However, as shown in Table 4.29, there was a significant difference between the two groups of words on the SPT (IPT) (Z = -2.134, p =0.03, d = 0.72), which indicates that high frequency of repetition did significantly enhance learning and retention of the target non-words in the IPT. This difference was no longer visible in the DTP.

Table 4.28 Mean accuracy scores on the VKS and standard deviations of target non-words appearing eight times or four times in the reading passages

N = 40	**VKS (Max = 4)**		
	NW occurring 8 times	**NW occurring 4 times**	**Sig/Diff**
Mean	2.90 72.50%	2.63 65.75%	NS
SD	1.26	1.29	

Table 4.29 Mean accuracy scores on the SPT (IPT and DPT) and standard deviations of the target non-words appearing eight times or four times in the reading passages

N = 40 (Max = 8)	**SPT (IPT)**			**SPT (DPT)**		
	NW occurring 8 times	**NW occurring 4 times**	**Sig/Diff**	**NW occurring 8 times**	**NW occurring 4 times**	**Sig/Diff**
Mean	6.34 79.25%	5.68 71.00%	0.03	6.84 85.50%	6.28 78.50%	NS
SD	1.98	2.37		1.77	2.41	

The speed with which learners processed non-words appearing eight times and four times was also analysed (see Table 4.30). For both the IPT and the DPT, the results revealed that there were no significant RT differences for non-words appearing with different frequencies. This means that the frequency of occurrence of the target non-words in passages A and B did not significantly explain the speed with which learners recognised the meaning of the non-words in the SPT.

Table 4.30 Means and standard deviations of the RTs for target non-words appearing eight times or four times on the SPT (IPT and DPT)

N = 40	SPT (IPT)			SPT (DPT)		
	NW occurring 8 times	**NW occurring 4 times**	**Sig/Diff**	**NW occurring 8 times**	**NW occurring 4 times**	**Sig/Diff**
Mean	912.16	872.86	NS	838.77	858.73	NS
SD	252.57	212.75		165.64	245.29	

The potential contribution of the frequency of the original words (which had been replaced by target non-words) to learning and retention of these words was examined. Two target non-words replaced high-frequency (1k) real words: *cenedies,* meaning "human" and *sataca*, meaning "trip," and two non-words replaced low-frequency real words: *toroko,* meaning "sting" (4k), and *pocoko,* meaning "suburb" (3k). It was also expected (Prediction 2 of RQ3) that target non-words which replaced high-frequency words in the text would be better learned and retained. There were no significant differences in accuracy between non-words which replaced high-frequency words (HF) and those which replaced low-frequency words (LF) on VKS and SPT (IPT and DPT), which suggested that word frequency of the original words did not contribute to learning and retention of the target non-words (see Tables 4.31 and 4.32).

Table 4.31 Mean accuracy scores and standard deviations on the VKS for target non-words which replaced high-frequency words (HF) and low-frequency words (LF) in the reading passages

VKS (Max = 4)			
N = 40	**HF - NW**	**LF - NW**	**Sig/Diff**
Mean	2.68 67.00%	2.85 71.25%	NS
SD	1.21	1.35	

Table 4.32 Mean accuracy scores and standard deviations on the SPT (IPT and DPT) for target non-words which replaced high-frequency words (HF) and low-frequency words (LF) in the reading passages

(Max = 8)		**SPT (IPT)**			**SPT (DPT)**		
		HF - NW	**LF - NW**	**Sig/Diff**	**HF- NW**	**LF - NW**	**Sig/Diff**
N = 40	Mean	5.87 73.38%	6.21 77.63%	NS	6.24 78.00%	6.89 86.13%	NS
	SD	2.26	2.25		2.34	1.97	

The focus was then shifted to the speed processing (RTs) of the original words from Passages A and B which had been replaced by target non-words. As can be seen in Table 4.33, in the IPT no significant difference was found for RTs for items which replaced high-frequency words (HF) and items which replaced low-frequency words (LF). In the DPT, by contrast, a significant difference between these two categories was revealed (Z = -2.679, p = 0.007, d = 0.94). Contrary to expectations, however, the speed with which respondents recognised these words was shorter for the LF items than for the HF items in the SPT (DPT).

Table 4.33 Mean and standard deviations of the RT of target non-words which replaced high-frequency words (HF) or low frequency (LF) on the SPT (IPT and DPT)

N = 40	**SPT (IPT)**			**SPT (DPT)**		
	HF - NW	**LF - NW**	**Sig/Diff**	**HF - NW**	**LF - NW**	**Sig/Diff**
Mean	919.22	875.45	NS	871.01	809.54	0.007
SD	229.86	259.50		217.17	167.62	

In summary, as expected, the SPT (IPT) results showed that target non-words which occurred eight times were better retained than target non-words which occurred four times. It was also

observed that word frequency of the words from the original text (which had been replaced by target non-words) did not explain the learning and retention of these non-words. The RT results revealed that the frequency of occurrence of the target non-words and the frequency of the original words which had been replaced by the target non-words did not significantly contribute to the speed with which learners processed the meaning of the target non-words.

4.5 Discussion

4.5.1 Introduction

The purpose of this section is to discuss the results of Study 2 described above. First of all, the theoretical implications are examined in the light of the available research included in the literature review chapter, showing the extent to which the present findings compare to those in the relevant literature. The last sub-section below presents some pedagogical implications of the findings that need to be considered when using L2 vocabulary activities.

4.5.2 Theoretical implications of the results

The three key theoretical implications arising from the results presented in Study 2 are summarised in Figure 4.3.

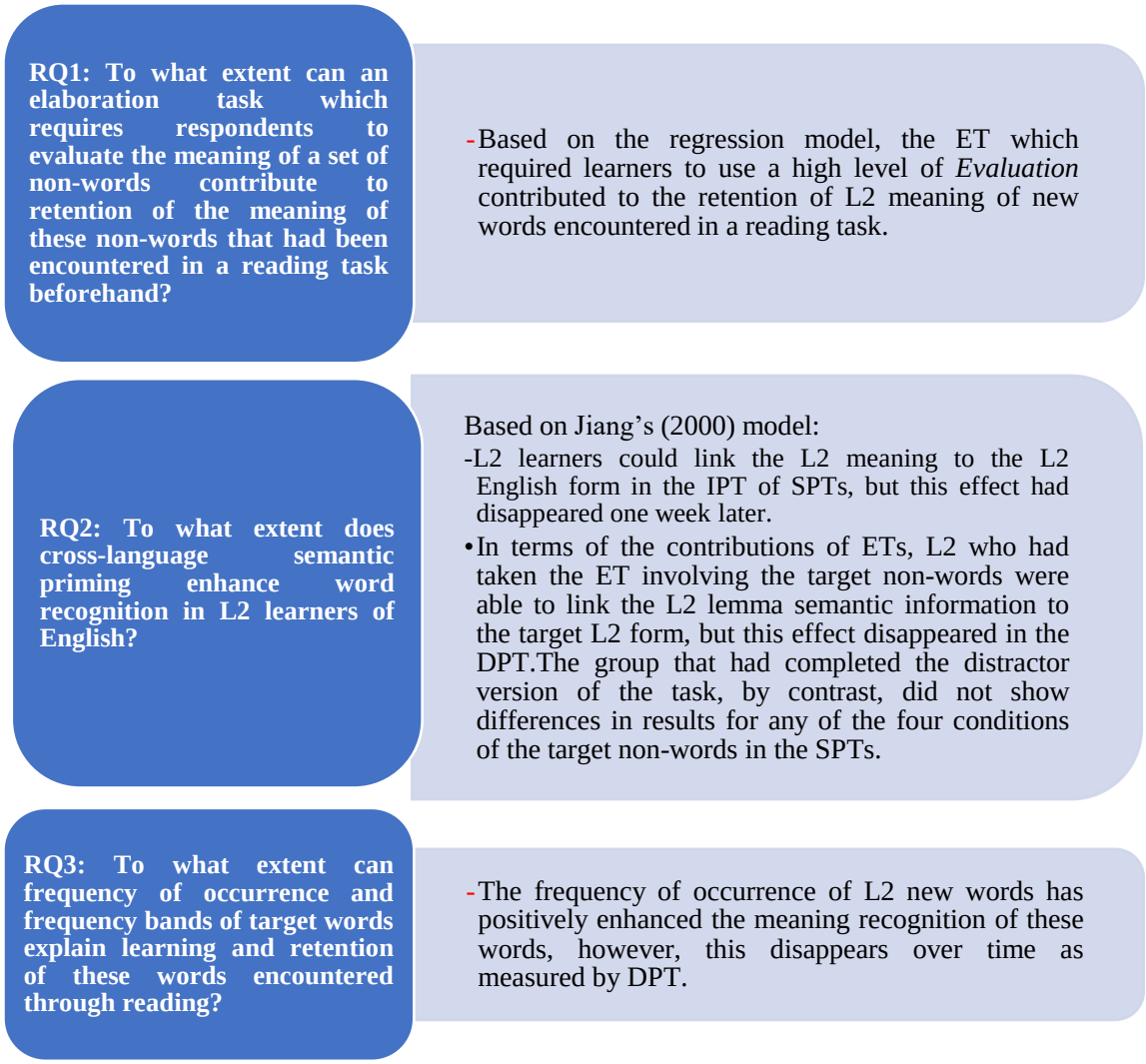


Figure 4.3 Key findings of Study 2

This study sought to investigate to what extent an elaboration task (ET) that requires participants to reflect on the meaning of a set of target words can enhance these learners' retention of the meaning of these words that had been encountered previously in a reading task. The results of this study show that an ET has a significant positive correlation with retaining new words, as measured by both immediate and delayed post-tests, confirming the possibility of incidental vocabulary learning through reading if relevant post-reading activities are undertaken.

Based on the regression model, the results show that ETs which required participants to reflect on non-words significantly contributed to participants' retention of target non-words, as shown by the offline and the receptive recognition online post-tests. This fully supports earlier research by Paribakht and Wesche (1997), who found that practising new words in a post authentic reading exercise can enhance learning these new words. The present results also provide evidence for the Involvement Load Hypothesis (ILH), as in Hulstijn and Laufer (2001), and indicate that ETs with

a high level of *evaluation* can contribute to furthering L2 vocabulary learning and retention. This finding contrasts with Folse (2006), who claims that depth of processing is not a factor in word learning and that the of word exposure alone can mainly explain word learning. Likewise, Yaqubi et al. (2010) found that L2 learners who completed different tasks with a lower involvement index obtained higher scores than L2 learners who carried out tasks with a higher index.

The current findings from the VKS accuracy scores reveal, as expected, that the meaning of target non-words was better retained in the group that had carried out ETs with these words. This also confirms Laufer and Hulstijn's (2001) viewpoint on the ILH and indicates that ETs with a higher degree of involvement result in better acquisition and retention of the meaning of new words than ETs with a lower degree of involvement. However, the contribution of the ETs could not be detected in the SPTs, as shown in the absence of differences in accuracy scores and RTs between groups taking ETs with or without target non-words. The main reason for this result may be due to factors such as time pressure, participants' familiarity with the task, and the possibility of answering correctly by chance, which should be considered when discussing the SPT data. The comparison between pre-tests and post-tests reveals that the participants' scores increased in the post-tests, which clearly suggests that the participants were more familiar with the format of the SPT in the DPT, and their familiarity with the SPT format can be considered as an important factor that may affect their responses. This contradictory finding clearly requires further investigation in order to find out to what extent a post-reading task with a high involvement component can contribute to L2 word retention and how this contribution is best measured.

The general percentage accuracy scores indicate that participants are better at retaining real words than target non-words, as was found in Study 1, which is to be expected, as learners are often more familiar with words they already know than with newly learned words. The learners' RTs for real words are also shorter than for target non-words in both the immediate and delayed post-test of the SPT. The results further reveal that the mean accuracy scores of L2 learners increase for both real words and target non-words from the IPT to the DPT. The RT results also show that, for both real words and non-words, the participants' processing speed in the DPT is shorter than in the IPT. On the one hand, as mentioned previously, this could mean that the participants were more familiar with the format of the SPT and therefore performed better the second time round. On the other hand, it is possible that the effects of the learning effects are only visible at a later stage, as suggested by Kim (2011), who notes that the impact of different involvement tasks may not be shown immediately but only at a later stage.

In terms of Jiang's (2000) three-stage model of L2 lexical development, it is perhaps possible to claim that students were able to map L2 semantic content onto the target L2 form in the IPT, but these L2 learners revert to relying more often on the L1 word meanings in the DPT. The L2 learners in the current study, thus, could link the L2 meaning to the L2 English form in the IPT of SPTs, but this effect disappeared one week later. This pattern confirms Jiang's (2000) suggestion that most of the L2 words become fossilised in the second stage (i.e., the L1 lemma mediation stage), due to the previously established L1 lemma (semantic information) mediation, which may prevent L2 learners from extracting L2 lemma information from the input. The difficulties in creating an L2 lemma may be due to the L1 lemma availability for those L2 learners, which may block access to L2 lexical entry. Moreover, factors other than input and the existence of L1 lemma information that were not studied in this project may have affected L2 lexical processing and development.

With respect to the contributions of the ETs, participants who carried out the ET involving non-words in the IPT were better at recognising target words after primes that were related to English and Arabic than at recognising target words that were shown after primes that were related only to Arabic, but this effect was reversed in the DPT. The RT results also indicate that participants who took the non-word ET continued to link the L2 lemma semantic information to the target L2 form. In contrast, the group that completed the distractor version of the task showed neither differences in any of the four conditions of the target non-words in the IPT or the DPT, nor any differences in processing times in either the IPT or the DPT. Therefore, these results support the conclusion that tasks with higher involvement load (such as the ET in this study) enhance vocabulary learning and retention in reading tasks. This again strongly supports Laufer and Hulstijn's (2001) claim, based on the ILH: tasks that require processing words more deeply result in better acquisition and retention of new words.

Non-words that replace HF words or occur more frequently in texts were also expected to be better retained. The results from the current study reveal that target non-words occurring more frequently are better retained in the IPT. These findings directly support studies on frequency of word exposure through reading (Horst et al., 1998; Pigada & Schmitt, 2006; Waring & Takaki, 2003), which provide evidence for the contribution of word frequency to incidental vocabulary learning. The present findings also concur with Study 1's conclusions that, for L2 learners, the enhancement of frequency disappears over time, beginning one week later. These findings also confirm Eckerth and Tavakoli's (2012) point that incidental vocabulary learning needs to be built up through repeated exposure, in addition to an elaborate processing of L2 lexical items. Contrary

to expectations, this study revealed that the frequency with which target non-words occurred did not explain learners' speed in processing the meaning of the non-words. The present findings further confirm Study 1's results, as the frequency of the original words that are replaced by target non-words did not increase learners' chances to learn or retain the meaning of new L2 words or enhance their speed in processing the meaning of new L2 words.

4.5.3 Pedagogical implications

The current findings show how reading tasks can enhance L2 vocabulary learning, which can have significant implications for L2 pedagogy. Instructors need to offer students opportunities to develop their vocabulary learning skills through reading by using enhancement activities. As discussed previously in Section 4.5.2, post-reading activities (such as the ETs in the current study) that include target words and require L2 learners to process the meaning of these new words in different ways can be a much more effective way to learn L2 vocabulary. Teachers, hence, should be conscious of the importance of formulating tasks based on the ILH. For example, instructors should be encouraged to design tasks with a high involvement load to enhance learning, including understanding different aspects of word knowledge through reading.

Time spent on tasks (Kim, 2008; Tang & Treffers-Daller, 2016) and the design of suitable materials for learning, such as the number of new words encountered in a reading task (Nation, 2013b), can further affect L2 vocabulary acquisition. Well-designed vocabulary tasks in which the proportion of unknown words in a text is appropriate for the learners are also needed in L2 classes. The results for the four conditions of the target non-words suggest that the semantic links between English and Arabic words should be considered in L2 vocabulary teaching and learning.

As Jiang (2000) suggests, L2 words must be taught in L2 contexts which provide repeated exposure to new words via reading in order to enhance learners' chances to acquire the meaning of new words in L2 contexts. This, then, could help decrease L2 learners' tendency to rely on their L1. The current findings confirm this approach by showing that target non-words that occur eight times are better retained than target non-words that occur four times. Teachers can also design reading tasks in which students can process target words several times in various L2 contexts. This approach can provide better conditions for L2 lexical competence development.

This study further demonstrates that learners are better at retaining target non-words in the DPT than in the IPT, which clearly shows the importance of DPTs, since the impact of learning new words may appear at a later stage. Finally, as the above discussion of the practical implications

of the present research has highlighted, an awareness of the importance of depth of processing can help teachers to develop tasks that are informed by theories of L2 incidental vocabulary learning.

Chapter 5: Conclusion

5.1 Introduction

This chapter concludes the present study. It first relates the contributions made to knowledge in the field of L2 vocabulary acquisition. It then gives a summary of the key findings, thereby contributing to the understanding of how L2 lexicon can be learned and developed incidentally through reading by L1 Arabic learners of English. At the end of the chapter, the limitations of the current study are considered, and some suggestions for future research are provided.

5.2 Contributions to Knowledge

In this part of the conclusion, the contributions to the state of knowledge in the field of L2 vocabulary acquisition from reading that the current project has attempted to achieve are highlighted. In light of the key findings presented from the current study, these contributions are discussed in the following sub-sections.

5.2.1 L2 vocabulary learning and retention from reading

Both of the studies conducted in this research were carefully designed to obtain insight into how new words can be learned incidentally from reading by L1 Arabic learners of English. The findings from both studies provide empirical evidence that L2 learners can learn and retain new words incidentally from reading. This finding is in line with other results that confirmed the possibility of learning vocabulary from reading (Waring & Takaki, 2003; Horst, 2005; Pigada & Schmitt, 2006; Brown et al., 2008). The first part of the current study (Study 1) showed that L2 learners could learn the form (receptive level) of about three-quarters (74%) of the target words, while they could learn the meaning of the target words and use them in a sentence (productive level) at a rate of 4.88%. This indicates that at the productive level, approximately one in twenty tested words was learned, while at the receptive level, three out of four tested words were learned incidentally from reading. Two weeks later, the participants were shown to have to retained more than half (63.50%) of the forms (receptive level) of the target words, meaning that 2.25 of every 4 tested words were incidentally retained from reading. In addition, the Study 2 findings showed that L2 learners were able to recognise the meaning of 70% of the target words (receptive knowledge) and learn the meaning and use (productive knowledge) of 69.06% of the target words in a sentence. This means that for both receptive and productive knowledge, approximately 2.75 of each 4 tested words were incidentally learned from reading.

The participants were also able to retain the meaning of 80.94% of the target words a week later, which means that 3.25 out of 4 words were retained. Thus, although the rates of learning and retention seemed to be quite low, particularly in Study 1, it can be argued that reading comprehension can lead to incidental vocabulary acquisition and enrich aspects of word knowledge.

5.2.2 Receptive and productive learning

The results of Study 1 indicated that reading enhanced incidental L2 vocabulary growth at the receptive level better than at the productive level of word knowledge for L1 Arabic learners of English. In contrast, the results of Study 2 showed that both the receptive and productive levels of word knowledge were enhanced by incidental reading. An important conclusion about these two levels of word knowledge, therefore, can be drawn from the present study: the results suggest that to learn the new words productively, L2 learners need to process these words at deeper levels and more elaborately. This can be accomplished by providing more elaborate processing activities that focus more on the meaning or elements of using the words in a sentence correctly. As shown in the Study 1 results, it seems that receptive knowledge of a word is learned first, and, as appeared in Study's 2 findings, encouraging L2 learners to engage more with the meaning of a word can then facilitate its learning at the level of productive knowledge. This is plausible, as L2 vocabulary learning seems to be a progressive process; as Melka (1997) pointed out, in this process, receptive word knowledge is normally more advanced and productive word knowledge is often learned later than receptive word knowledge.

5.2.3 Retention rates in incidental L2 vocabulary learning

While this study provided evidence about the amount of new vocabulary that can be learned incidentally from reading, it also gave insight into how the retention rates of incidental vocabulary increase over time. The delayed post-test in Study 1 showed that L2 learners could retain their newly learned words, and this effect persisted for nearly two weeks. However, as suggested by Waring and Takaki (2003), new words learned from reading are likely to be forgotten quickly, and they completely vanish from memory within a few months. However, the results of Study 2 revealed that L2 learners' target word scores increased a week later, and the RT results showed that for the new words, the participants' processing speed in the test administered 1 week later was shorter than in the immediate test. This

suggests that processing new words more deeply is beneficial for long-term retention of the new words learned incidentally from reading.

5.2.4 L2 lexical development in a reading task

Study 1 revealed that the forms of new words are learned and retained incidentally from reading better than their meanings or use. This is in line with earlier findings suggested by Jiang's (2000) model of vocabulary learning among L2 learners, namely, that the form of the L2 lexeme can be learned at an earlier stage, while L2 meaning and use is integrated into L2 lexical entries at later stages. However, after processing these new words more deeply in Study 2, the results reveal that the L2 lemma semantic information seems to be mapped into the target L2 form in the immediate post-test, but these L2 learners revert to relying more often on the L1 word meanings one week later. This clearly suggests the importance of increased L2 learner engagement with the new words they have learned, which will help them to link the L2 lemma content to the L2 form, and, as a result, start building a complete L2 lexical entry in L2. However, as shown in Study 2, this effect is likely to disappear within a few days. This replicates the pattern previously proposed in Jiang's (2000) model, where most L2 words become fossilised in the second stage (i.e., the L1 lemma mediation stage). This shows that new words need to be processed more deeply to increase the chances that L2 lexical entries can be developed, as suggested by Jiang's model, where new semantic information from the L2 can be integrated into the lemma.

5.2.5 The contribution of frequency in L2 vocabulary learning and retention

The immediate post-test in Study 1 showed that word frequency of occurrence enhances L2 learners' ability to recognise the form of the target words (receptive knowledge). Study 2 supported this evidence, as shown in the immediate test, as new words occurring more frequently were retained better than those that occurred less frequently (receptive knowledge). This finding, therefore, confirms the importance of repeated exposure to new words when it comes to enhancing incidental L2 vocabulary learning and clearly suggests that the more words are repeated in a text, the better chance there is for receptive knowledge to be obtained.

The present findings showed that to enhance receptive gains, L2 learners need to encounter the new words at least eight times, and they may need further exposure to the new words to enhance their incidental vocabulary learning at a productive level. These findings directly support previous results on the frequency of word exposure needed in reading (Saragi et al., 1978; Webb, 2007), providing evidence for the significant enhancement of word frequency on

incidental receptive learning. However, the delayed post-test in Study 2 supported Study 1's finding that the contribution of frequency disappears over time for L2 learners, which again indicates the importance of increased L2 learner engagement with the new words they have learned for long-term retention. This suggests a new insight that the frequency of word exposure is a necessary but not a sufficient condition to enhance incidental L2 vocabulary learning and retention. Hence, it can be argued that in addition to the importance of repeated exposure, more elaborate processing of the new words is also essential, where gaining receptive knowledge can be a step towards obtaining productive word knowledge, and then, as shown in Section 5.2.3, processing the words more deeply can enhance the long-term retention of the new words learned incidentally from reading.

5.2.6 The contribution of vocabulary depth to reading comprehension

One of the key findings in Study 1 was that in summarising the reading stories, only the aspects of meaning and use at productive level of word knowledge significantly explained variance in reading comprehension (16% of the variance in reading comprehension). This illustrates an important point, namely, that depth of vocabulary at a productive level is more important for reading comprehension performance than vocabulary size (Rashidi & Khosravi, 2010). This may justify the assumption that learners who are better at understanding words in depth are also better at reading comprehension performance.

The data analysis from Study 1 also revealed a strong correlation between summarising the reading text and incidental vocabulary learning at a productive level of knowledge. This indicates that learners who are good at summarising stories are also proficient in producing the meaning and use of new words. In other words, L2 learners who have a greater depth of word knowledge can be more confident in determining and guessing the meaning of new words, which leads to success in reading comprehension. This is in line with a previous finding concerning the importance of depth of word knowledge for reading comprehension (Zahang &Yang, 2016) and confirms that enhancing the depth of vocabulary knowledge (aspects of meaning and use in a sentence) at a productive level for specific words can contribute to successful reading.

5.2.7 The role of the Involvement Load Hypothesis in L2 vocabulary learning and retention

The findings of Study 2 provide empirical evidence for the claim that the more L2 learners process and engage with new words, the better they will learn and retain them. Moreover, the results of this study revealed that the post-reading task, which required L2 learners to engage with the new words in the text more deeply by focussing on the meaning and

the way these words are used grammatically, demonstrated significant enhancement in their incidental L2 vocabulary learning at both the receptive (recognising L2 meaning) and productive (providing L2 meaning and use in a sentence) levels of knowledge. In contrast, Study 1 indicated that the form of the new words at the receptive level was better learned and retained incidentally from reading than the meaning or use at the productive level, for which an extremely low rate was observed. This means that receptive knowledge of a word is learned first, and depth of processing levels for the word, as shown in Study 2, may then facilitate its learning at the level of productive knowledge.

As discussed above, the Study 1 results showed that there were no differences in L2 learners' performance between their immediate and delayed post-test, which means that they retained the newly learned words; this effect persisted for nearly 2 weeks. In contrast, the Study 2 results showed that L2 learners' target word scores increased a week later, and the RT results also reveal that for the new words learned, the participants' processing speed in the test administered 1 week later was shorter than that in the immediate test. This clearly suggests that processing the new words more deeply is crucial for long-term retention of new words learned incidentally from reading.

In terms of L2 lexical development, Study 2's immediate findings appeared to reveal that L2 learners could link the L2 meaning to the L2 English form in the IPT of the SPTs. But the delayed test showed that L2 learners rely more on the Arabic semantic representation of the target words' meaning which may justify the assumption that L2 learners continue to map the L1 lemma information onto the L2 form a week later. The delayed post-test in Study 2 and the findings from Study 1 further showed that, for L2 learners, the effects of word exposure frequency disappear over time, which again suggests the value of continuing L2 learners' depth of processing with the new words to facilitate long-term retention. Thus, it can be argued that if reading is to contribute to incidental L2 vocabulary learning, this is only possible if more elaborate processing activities are provided in which L2 learners engage with the meaning of the word in the text.

5.3 Contribution to knowledge: Innovation in research design

The present study provides a further contribution to the field of incidental L2 vocabulary learning research in terms of the research design, as described below.

5.3.1 The experimental design

To achieve the aims of the present project, an innovative experimental design was adopted. The design of the current project was carefully created to overcome some problematic issues described in previous studies. In the current study, the L2 target non-words were created according to Arabic rules to suit Arabic learners of English. In addition, the density of unknown words in the text was carefully controlled, based on previous findings (e.g., Laufer & Sim, 1985; Laufer, 1989; Hsueh-chao & Nation, 2000; Schmitt et al., 2011).

The participants who were involved in Study 1 were selected from Qassim University in Saudi Arabia several weeks before the experiment was carried out. The selection of the passages was based on the scores on the SWRT, which was conducted with L2 learners to assess their ability to decode words and their sight-reading skills in order to select the appropriate level of difficulty of the YARC passage reading test for them. After selecting the appropriate passages, the researcher manipulated the original text of the YARC to ensure that reading the text would be challenging but not too complicated. Thus, the two reading texts were carefully controlled to prevent the inclusion of a large number of unknown words (2.5% of the text was composed of target non-words) encountered in each passage through two different exposures. This study also confirmed the findings from Waring and Takaki (2003) and Brown et al. (2008) that replacing real words in the texts with non- words as target words is an appropriate method to ensure that the participants were not familiar with them in advance. This design also added further value to the current project in terms of the way it considered the difficulty levels of the texts and measured the learners' actual vocabulary gain from reading.

It has been mentioned in the literature that none of the separate word knowledge tests can completely assess the whole concept, which requires more than one measurement for more precise assessment (Li & Kirby, 2014). In both of the present studies (Study 1 and Study 2), therefore, a combination of vocabulary measure formats, comprising offline and online recognition tests, were employed to assess different aspects of word knowledge. This is an important strategy allowing for more accurate data and results. By using the online software tests adopted in this research, it was possible to measure lexical access and word recognition processes in detail. More importantly, this allowed the researcher to measure the speed of the participants' responses, thereby reflecting how well they could access and recognise the word. Finally, it is hoped that the methodology adopted in this research can be a useful example that may encourage other researchers to try out similar experimental designs.

5.4 Limitations of the study

Although considerable empirical support was found for the claims made in the current study, there are several notable limitations related to different features of the experimental design, and these should be highlighted and explained. These limitations are discussed in the following sub-sections.

5.4.1 Sampling

The number of participants in Study 1 per group (G1 [N = 30], G2 [N = 28]) and particularly in the LDT results (G2 [N = 24]) was slightly low. In Study 2, there was a total of 40 participants (G1 [N = 20], G2 [N = 20]). Therefore, it is possible that a Type II error has occurred, because of lack of statistical power. A power analysis accredited by the Clinical and Translational Science Institute at the University of California, San Francisco was then conducted to estimate the sample size needed. The online sample size calculator shows that if the effect size was expected to be between medium to strong, then between 25 to 63 participants would be required per group. This confirms that the number of participants was rather limited for a study of this kind.

Two main reasons for not involving a larger number of Arabic ESL learners in this research will be considered here. The first one is related to the complexity of the current experimental design for both Study 1 and Study 2; because of this complexity, many L2 learners were not willing to participate in this research. As discussed in Chapter 1 (Section 1.4), due to L2 learners' limited experience in reading practice in English and deficiencies in their English vocabulary, most of the students opted not to volunteer because they felt that it would be too challenging for them to complete all the required tasks in the research. The second reason is related to the challenging situation in the university at the time this project was conducted. The researcher's data collection time was extremely limited, and most of the students were busy preparing their materials and research for the mid-term examinations, and ultimately, their final examinations. The researcher made extensive efforts to manage these challenges, complete the experimental design and follow all procedures required for both Study 1 and Study 2 with the greatest possible number of L2 participants.

5.4.2 A new measurement to assess L1 Arabic

The YARC test was designed for native English speakers aged 11-16 years, and, to the author's knowledge, no other studies have used the YARC test with a sample of postgraduate L1 Arabic learners of English. Although conducting the YARC test in the present research may be

considered a contribution to knowledge in the field, it may also be considered as one limitation of the current study's design because researchers have yet to confirm the YARC test's reliability and validity in the adult Arabic ESL context. In order to overcome this issue, as a first step, the topic of YARC passages was carefully selected to be consistent with the Arabic context. The first passage discussed a bird called the "Honey Guide," and the second passage was an interesting short story about a "River Girl," that is, both selected topics were well known and common with regard to L2 participants' background. In the second step, the features of each text were controlled, as previously mentioned (Section 3.3.5.4), by investigating the level of difficulty of the reading passages using *the vocabulary profiler* (BNC-20 version), available on *the Compleat Lexical Tutor* website (Cobb, n.d.) to ensure that the participants would be familiar with all words in the text except the target non-words. The suitability of YARC Secondary Test for adult learners of English was further identified, and it was found, in terms of L2 participants' SWRT scores, that L1 Arabic adult students fell within the range for native English children. This way, it seems that the YARC Secondary Test can probably be considered as suitable to be completed by the L1 Arabic learners of English who participated in the current study.

5.4.3 Two texts including a limited number of target words

One of the limitations in the current study is that building the whole study on a limited number of target non-words may not be enough to draw strong conclusions from the results. Indeed, using a small set of non-words as targets may lead to learning them easily, and then they can be simply retained. This may affect the validity of some of the statistics reported in the results. In other words, the rate of learning and retention reported in this study could be changed if a larger number of target non-words were used. This limitation was beyond the researcher's control because involving a larger number of target non-words required the participants to read a larger amount of text or more than two texts. This was difficult to achieve, because, as previously mentioned, the data collection time was very limited and most of the L2 learners were not motivated to complete a long reading task. Therefore, replication of the present study is suggested using more texts, and then more target non-words for more generalisable findings.

5.4.4 Priming condition result

After analysing Study 2's findings, an issue regarding the results of the priming conditions has been observed. The SPT results revealed that in terms of accuracy scores, there was evidence in the IPT that L2 learners could link the L2 meaning to the L2 English, but this effect disappeared in the DPT as they were relying on the Arabic representation of the target. In contrast, with

respect to RTs, no effect was found in the IPT, but there was evidence that participants had started to build lemmas with L2 word semantics in the DPT. It can be noticed that accuracy score results share the RTs, an important finding which is consistent with Jiang's (2002) view, suggesting the existence of L1 semantic concept in L2 lexical entries. However, this finding of the current study differed in suggesting when precisely L2 learners had started to map L2 meaning to L2 words. This contrasting finding indicates, as claimed by (Kaufman, Deyoung, Gray, Jimenez, Brown, & Mackintosh, 2010), that RT measures may not necessarily represent learning ability, due to the possibility of providing unstable data which may not be sufficiently reflective of learning. For instance, learning cannot be determined only by the total reduction in RTs over successive blocks of trials, due to practice or being more familiar with the task format. Therefore, the current pattern of results suggested by the accuracy scores seems to provide strong evidence and a more accurate reflection of the connection between lemma and L2 semantic information. Although the RTs do not seem to affect the main finding (the accuracy score result), nevertheless, this may limit the current results, and requires a clear understanding of the connections between L2 words and conceptual representation by demanding further investigation into the process of L2 lexical representation and development.

5.4.5 Generalisability

First, since all L2 learners included in Study 1 were students at one university, it can be assumed that the results cannot necessarily be generalised to other universities. Although it is expected that other L1 Arabic ESL learners who belong to other universities in Saudi Arabia are involved in a similar education system to those at the target university, the findings from this project would have been generalisable to other educational contexts if it had included participants from other institutions. To overcome this issue, L1 Arabic learners of English who participated in Study 2 were all postgraduate students at Reading University, and they had graduated from different universities within and outside Saudi Arabia, Libya, Yemen, and Qatar.

5.4.6 Vocabulary learning rates represented by the immediate post-test

In relation to the procedures of the current research design, the immediate post-test represented learners' incidental vocabulary learning rates; however, this test was given to the participants immediately after they had completed reading. It could be argued, as Schmitt (2010) suggested, that learning reported from the immediate post-test does not represent a full picture of the actual acquisition, as the knowledge learned can decrease over time. Another issue in relation to the immediate post-test, is that it can also be observed that such a test can attract

participants' attention to vocabulary learning, making it difficult to assume that their vocabulary learning was completely incidental. Hence, the researcher in the present study was extremely careful when communicating with the participants and refrained from using any words that could indicate that they would be tested or would have to complete a vocabulary test, in order to keep their attention away from the aims of the study.

5.4.7. Further post-test required

As the present study investigated incidental L2 vocabulary learning and retention from reading over a short period – a maximum of two weeks in Study 1 and only one week in Study 2 – it would be useful to administer another post-test after a longer period to determine how long L2 learners retain words learned through reading in the long term. This would be especially helpful in gaining a clearer picture of the depth of word knowledge related to L2 lexical development, as it was found in Study 2 that processing new words more deeply is required for long-term retention of new words learned incidentally from reading.

5.5 Suggestions for further research

This study sought to investigate three aspects of word knowledge (form, meaning, and use in a sentence) in relation to reading comprehension by Arabic ESL students. Thus, in terms of expanding the scope of the present experimental research, further studies could be carried out to evaluate the relationship between L2 learners' learning of other aspects of word knowledge and reading comprehension performance. This may be required to include various depth of word knowledge measurements with a view to gaining a more accurate assessment and a clear picture of how each aspect of L2 lexical knowledge can be developed incidentally through reading. Furthermore, as the present findings show that to enhance incidental receptive vocabulary learning, L2 learners need to encounter the new words at least eight times, it is suggested that further research should focus on additional exposure to the new words to enhance the learning of other aspects of word knowledge at a productive level. As Schmitt (2010) pointed out, more exposure to a word can lead to the gradual learning of other aspects of word knowledge, which obviously suggests directions for future studies to help learners develop a wider grasp of L2 word knowledge.

Following the methodology used by Waring and Takaki (2003), and Brown et al. (2008), the present study involves the insertion of non-words as targets into reading passages and certainly suggests the advantage of using these non-words in future research. Using such non-words as targets allows the researcher to ensure that participants would be totally unfamiliar with these

words prior to reading, and, further, they would not be exposed to them in between the immediate and delayed post-test. The researcher further can be sure that participants' background knowledge has no effect on their scores on the tests conducted, and then has no effect on their learning and retention of any aspects of word knowledge from reading, and no aspect had an advantage over the others as a result of participants' pre-existing knowledge. It would also be beneficial to ensure that participants' scores on the post-test after the delay would not be affected by their regular language studies. For these benefits, employing non-words as targets is recommended for future research.

The aims of the current project were achieved, but at the same time, the findings lead to new research questions that can be taken up in future L2 vocabulary acquisition research. They are listed as the following:

- As shown in Study 1, L1 Arabic learners of English seem to be equivalent with native English speakers as regards their learning and retention of new words, which clearly indicates the advantages of learning and retaining new L2 words. This explanation of the performance of L1 Arabic learners of English in the present study can be tested in future research by investigating to what extent the bilingualism can facilitate word learning from reading. This finding further suggests the importance of exploring the factors that may yield this bilingual advantage for learning new words from reading.

- The present results of Study 1 revealed that there was no significant relationship between size of vocabulary knowledge (vocabulary size) and reading comprehension performance, and then showed that L2 learners' vocabulary size does not contribute to successful reading. Hence, further studies need to consider the question of ensuring a suitable level of difficulty of a reading text with more advanced Arabic ESL learners. In other words, more advanced L2 learners would have a larger vocabulary size, and this could allow them to use more advanced and effective strategies that could help to enhance their reading comprehension performance.
- Study 2 showed a contradictory result as regards the absence of significant differences on the SPTs between groups taking ETs with or without target non-words and the results suggested by the regression model, showing that ETs which require participants to reflect on non-words significantly contribute to participants' retention of target non-words. Although the results provided by the regression model seem to provide strong evidence for the association of the ET with the retention of the target non-words, this

contradictory finding clearly requires further research to explore to what extent a post-reading task with a high involvement component can contribute to L2 vocabulary retention and how this contribution is best measured.

- As mentioned in Chapter 2 (Sections 2.11.1, 2.11.2), the more relevant skills for the current study were taken into consideration in the assessment of reading comprehension. However, there are other factors which may contribute to the current participants' reading comprehension performance, but are not controlled in the present research. This obviously demands a follow-up study which focuses more on reading processes. In such a study, other factors such as L1 reading ability that may contribute to L2 reading comprehension performance could be investigated.
- The regression analysis of Study 2 suggests that the ET which requires learners to use a high level of *Evaluation* contributes to the retention of L2 meaning of new words encountered in a reading task. However, since the multiple-choice comprehension questions were designed to be relevant to the target non-words, and the participants needed to know the meaning of the non-words in order to complete the task, it seems that reading comprehension (inferencing ability) may further contribute to the retention of the meaning of L2 words. Therefore, future research can be conducted focusing more on how learners deal with the target words in a reading task and considering other conditions that may enhance their vocabulary retention process.

- As shown in Study 2, processing new words at deeper and more elaborate levels can lead to better learning and retention of the meaning and use of these words. Future research should include more vocabulary tasks that can encourage L2 learners to engage more deeply with the new words; this may help to expand L2 learners' knowledge of these words and enhance their learning of other aspects of word knowledge.
- Finally, in relation to L2 lexical development through reading, based on Jiang's (2000) three-stage model, it is clear that future study is needed because different word categories may follow different patterns of development.

5.6 Lessons learned

During my doctoral research journey, I have developed many research skills, including skills related to the preparation, collection, analysis and presentation of research data. The research classes that were provided in my first year of PhD study were also useful and

valuable, as they facilitated my development, both in terms of these research skills and the critical evaluation competencies that were essential for my research. Furthermore, I have learned to be a more independent student, and, consequently, I have become more acquainted with the literature knowledge concerning the field of my research area. As I have improved my understanding of some major issues related to L2 learning vocabulary through reading, I have been able to enrich my knowledge of both reading comprehension and L2 vocabulary acquisition, reaching a deep understanding of these topics.

At the same time, my doctoral studies have presented unexpected challenges and difficulties. For instance, designing the experiments with the software SuperLab was not easy, and the situation concerning Qassim University in Saudi Arabia as Study 1 was being conducted created some challenges in terms of the research data collection within the limited planned time; analysis of the research data was also complicated, particularly when it came to Study 2, due to the complicated design of this experimental study. However, by working hard and managing time effectively, these problems were resolved. It is hoped that the intervention study presented here will inform the development of new strategies used by Arabic ESL learners to expand their L2 vocabulary knowledge and improve the teaching methods used to teach reading comprehension and L2 vocabulary learning in this context.

References

Abo-Rabia, S. (2002). Reading in a root-based-morphology language: The case of Arabic.*Journal of Research in Reading, 25*(3), 299-309.

Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adolf, S., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19*(9), 933-958.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425-438.

Akasha, O. (2013). Exploring the challenges facing Arabic-Speaking ESL students & teachers in Middle School. *ELT and Applied Linguistics, 1*(1), 12-31.

Alderson, J. C. (2000). *Assessing reading.* Cambridge, England: Cambridge University Press.

Alsaif, A. (2011). *Investigating vocabulary input and explaining vocabulary uptake among EFL learners in Saudi Arabia.* Unpublished PhD thesis, Swansea University.

Anderson, R. C., & Freebody, P. (1981). *Reading comprehension and the assessment and acquisition of word knowledge*. Champaign, IL: Center for the Study of Reading, University of Illinois at Urbana-Champaign.

Anthony, H. M., Pearson, P. D., & Raphael, T. E. (1993). Reading comprehension: A selected review. In L. M. Cleary & M. D. Linn (Eds.), *Linguistics for teachers*. New York, NY: McGraw Hill**,** pp. 282-307.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* New York. NY: Academic Press.

Baddeley, A. (1997). *Human memory: Theory and practice* (Rev. ed.). Hove, UK: Psychology Press.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review, 85*, 139-152.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*. pp. 47-89. New York, NY: Academic Press.

Baleghizadeh, S. & Abbasi, M. (2013). The effect of four different types of involvement indices on vocabulary learning and retention of EFL learners. *The Journal of Teaching Language Skills (JTLS), 5*(2), 1-26.

Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly, 43*(1), 79-103.

Bauer, T. (1996). Arabic writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 559-564). New York, NY: Oxford University Press.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101-118.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues*. *Memory in everyday life*, 1, pp. 396-401.

Bogaards, P. (2000). Testing L2 vocabulary at a high level: The case of the Euralex French tests. *Applied Linguistics, 21,* 490-516.

Braze, D., Katz, L., Magnuson, J. S., Mencl, W. E., Tabor, W., Van Dyke, J. A., Gong, T., Johns, C. L., & Shankweiler, D. P. (2016). Vocabulary does not complicate the simple view of reading. *Reading and Writing, 29*(3), 435-451.

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities, 40,* 226-243.

Brown, C., Culligan, B., & Phillips, J. (2013). *A New General Service List (1.01)*. Retrieved from http://www.newgeneralservicelist.org

Brown, D. (2008). Using a modified version of the Vocabulary Knowledge Scale to aid vocabulary development. *The Language Teacher, 32*(12), 84-105.

Brown, H. C. (2004). *Language assessment: Principles and classroom practices*. New York, NY: Pearson Education.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading while listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136-163.

Bruton, A. (2009). The Vocabulary Knowledge Scale: A critical analysis. *Language Assessment Quarterly, 6* (4), 288-297.

Burnett, J. D. (1974). Parallel measurements and the Spearman-Brown formula. *Educational and Psychological Measurement, 34*(4), 785-788.

Carlisle, J. F. (2003). Morphology matters in learning to read: A commentary. *Reading Psychology, 24*, 291-322.

Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics*, 12 (2), 159-179.

Carrell, P. L., & Grabe, W. (2002). Reading. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 233-250). London, England: Edward Arnold.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354-380.

Chall, J. S. (1987). *Two vocabularies for reading: Recognition and meaning*. In M. G. McKeown & M. E. Curtis (Eds.). *The nature of vocabulary acquisition* (pp. 7-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693-713.

Chen, R. S., & Vellutino, F. R. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research, 29*(1), 1-24.

Chincotta, D., & Hoosain, R. (1995). Reading rate, articulatory suppression, and bilingual digit span. *European Journal of Cognitive Psychology, 7*, 201-211.

Cho, K. S., & Krashen, S. D. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, *37*(8) 662-667.

Clarke, M. A. (1979). Reading in Spanish and English: Evidence from adult ESL students. *Language Learning, 29,* 121-50.

Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition*. Cambridge, England: Cambridge University Press.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*, 301-315.

Cobb, T. (2010). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning, 12*(4), 345-360.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-238.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294.

Cui, Y. (2009). The development of lexical organization in Chinese EFL learners at tertiary level. *Arizona Working Papers in SLA & Teaching, 16*, 57-73.

Curtis, M. E. (2002). *Adolescent reading: A synthesis of research.* Paper presented at the National Institute for Literacy/National Institute of Child Health and Human Development Adolescent Literacy Workshop II, Baltimore, MD.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10,* 277-299.

Cziko, G. A. (1980). Language competence and reading strategies: A comparison of first and second-language oral reading errors. *Language Learning, 30,* 101-14.

Dallar, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modeling and assessing vocabulary knowledge*. Cambridge, England: Cambridge University Press.

Day, R., & Bamford, J. (1998). *Extensive reading in second language classroom.* Cambridge, England: Cambridge University Press.

Day, R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language, 7,* 541-551.

De Ramírez, R. D., & Shapiro, E. S. (2007). Cross-language relationship between Spanish and English oral reading fluency among Spanish- speaking English language learners in bilingual education classrooms. *Psychology of Schools, 44,* 795–806.

Devine, J. (1987). General language competence and adult second language reading. In J. Devine, P. L. Carrell, & D. E. Eskey (Eds.), *Research in reading in English as a second language.* Washington, DC: TESOL.

Doest, L., & Semin, G. (2005). Retrieval contexts and the concreteness effect: Dissociations in memory for concrete and abstract words. *European Journal of Cognitive Psychology, 17*(6), 859-881.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first and second-language learners. *Reading Research Quarterly, 38*(1), 78-103.

Dupuy, B., & Krashen, S. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning, 4*(1), 55-63.

Durgunoglu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85*, 453-465. http://dx.doi.org/10.1037/0022-0663.85.3.453

Ebbinghaus, H. (1885). Uber das Gedächtnis. Untersuchungen zur experimentellen Psychologie. Leipzig, Germany: Duncker & Humblot.

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research, 16*(2), 227-252.

Ehri, L. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9,* 167-188.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61,* 367-413.

Elley, W. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning, 41*(3), 375-411.

Ellis, N. C. (1994a). Implicit and explicit language learning: An overview. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp 1-32). London, England: Academic Press.

Ellis, N. C. (1994b). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. In J. H. Hulstijn & R. Schmidt (Eds.), *Consciousness in second language learning. AILA Review* II, 37-56.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L1 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194.

Engku Ibrahim, Engku Haliza, Isarji, S. arudin & Ainon Jariah, Muhamad. (2016). The relationship between vocabulary size and reading comprehension of ESL Learners. *English Language Teaching, 9*(2), 116-123.

Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics, 24,* 289-315.

Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language, 20*(1). Retrieved from http://www.nflrc.hawaii.edu/rfl/April2008/fender/fender.html

Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 116-132). Cambridge, England: Cambridge University Press.

Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*(3), 323-330.

Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*(2), 273-293.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680-690.

Forster, K. I., & Veres, C. (1998). The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning Memory and Cognition, 24*, 498-514.

Gardner, M. K., Rothkopf, E. Z., Lapan, R., & Lafferty, T. (1987). The word frequency effect in lexical decision: Finding a frequency-based component. *Memory & Cognition, 15,* 24-28.

Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition, 21*(2), 319-333.

Gathercole, S., & Alloway, T. (2007). *Understanding working memory: A classroom guide.* The University of York.

Gathercole, V. (2013a). *Issues in the assessment of bilinguals*. Bristol, UK: Multilingual Matters.

Gathercole, V. (2013b). *Solutions for the assessment of bilinguals*. Bristol, UK: Multilingual Matters.

Gee, N. R., Nelson, D. L., & Krawczyk, D. (1999). Is the concreteness effect a result of underlying network interconnectivity? *Journal of Memory and Language, 40,* 479-497.

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

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 10,* 341-363.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly, 25*(3), 375-406.

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

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading*. New York, NY: Pearson Education.

Greidanus, T., Beks, B., & Wakely, R. (2005). Testing the development of French word knowledge by advanced Dutch and English-speaking learners and native speakers. *Modern Language Journal, 89,* 221-233.

Guo, T., & Peng, D. (2005). Was the native *language* activated during the production of the second language: An ERP study with unskilled Chinese-English bilinguals. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society; New York.

Guo, Y., & Roehrig, A. D., (2011). Roles of general versus second language (L2) knowledge in L2 reading comprehension. *Reading in a Foreign Language, 23*(1), 42.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology, 56,* 1213-1236.

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund, Sweden: Lund University, Media-Tryck.

Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics, 10*(2), 221-240.

Haladyna, T. M. & Downing, S. M. (2004). Construct irrelevant variance in high stakes testing. *Educational Measurement: Issues and Practice, 23*(1), 17-27.

Hennebry, M., Rogers, V., Macaro, E., & Murphy, V. (2013). Direct teaching of vocabulary after listening: Is it worth the effort and what method is best? *Language Learning Journal, 1,* 19. http://dx.doi.org/10.1080/09571736.2013.84975

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303-317.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching, 41*(2), 87-106.

Hong, X. (2010). Review of effects of glosses on incidental vocabulary learning and reading comprehension. *Chinese Journal of Applied Linguistics, 33*(1), 56-73.

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

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review, 61*(3), 355-382.

Horst, M., Cobb T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106,* 501-511.

Hsueh-chao, M., & Nation, P. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Huang, J. (2013). *The Validity of the TEM-4 as a measure of vocabulary knowledge and reading comprehension.* (Master Dissertation). Institute of Education, University of Reading.

Huang, S., Eslami, Z., & Willson, V. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal, 96*(4), 544-557.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21*(2), 181-193.

Hughes, A. (1989). *Testing for language teachers*. Cambridge, England: Cambridge University Press.

Hughes, A. (2003). *Testing for language teachers*. Cambridge, England: Cambridge University Press.

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles & Coltheart. *Scientific Studies of Reading, 9*, 351-365.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud, & H. Béjoint (Eds.). *Vocabulary and Applied Linguistics*, (113-125). Basingstoke: Macmillan.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*(3), 327-339.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.

Hulstijn, J. H., & Trompetter, P. (1998). Incidental learning of second language vocabulary in computer-assisted reading and writing tasks. In D. Al-brechtsen, B. Henriksen, I. M. Mees, & E. Poulsen (Eds.), *Perspectives on foreign and second language pedagogy* (pp. 191-200). Odense, Denmark: Odense University Press.

Hunston, S., Francis, G., & Manning, E. (1997). Grammar and vocabulary: Showing the connections. *ELT Journal, 51*(3), 208-216.

Hunt, A., & Beglar, D. (1998). Current research and practice in teaching vocabulary. *The Language Teacher, 22(1)*, 7-25.

Jahangiri, K., & Abilipour, I. (2014). Effects of collaboration and exercise type on incidental vocabulary learning: Evidence against involvement load hypothesis. *Procedia-Social and Behavioral Sciences, 98*, 704-712.

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

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47-77.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *SSLA, 24*, 617-637.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, *88*(3), 416-432.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. Routledge.

Jiang, X. (2011). The role of first language literacy and second language proficiency in second language reading comprehension. *The Reading Matrix, 11*(2), 177-190.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19,* 357-377.

Johns, A. (1981). Necessary English: A faculty survey. *TESOL Quarterly, 15,* 51-57.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21*(2), 85-97.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Rev*iew, *16,* 705-710.

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The Involvement Load Hypothesis on trial. *Language Teaching Research, 12*(3), 365-386.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281-300.

Kieffer, M. J., & Lesaux, N. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing: An Interdisciplinary Journal, 21*(8), 783-804.

Kim, Y. J. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58(2),* 285-325.

Kim, Y. J. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 61(Suppl. 1),* 100-140.

Kirkpatrick, E. A. (1891). The number of words in an ordinary vocabulary. *Science, 18*(446), 107-108.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal, 78,* 285-99.

Koda, K. (2004). *Insights into second language reading: A cross-linguistic approach.* New York, NY: Cambridge University Press.

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.

Krashen, S. (1985). *The input hypothesis: Issues and implications.* New York, NY: Longman.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal, 73*(4), 440-464.

Krashen, S. (2004). *The power of reading*. Westport, Conn: Libraries Unlimited. Portsmouth, NH: Heinemann.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representation. *Journal of Memory and Language, 33,* 149-174.

Kudo, Y. (2001). Measuring L2 vocabulary. *A Focus on Language Test Development: Expanding the Language Proficiency Construct Across a Variety of Tests*, (21), 75.

Kuo, L. J., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist, 41*(3), 161-180.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). New York, NY: Academic Press.

Langeluddecke, P. M., & Lucas, S. K. (2003). Quantitative measures of memory malingering on the Wechsler Memory Scale (3rd ed.) in mild head injury litigants. *Archives of Clinical Neuropsychology, 18,* 181-197.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans to thinking machines* (pp. 316-323). Clevedon, UK: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Be´joint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London, England: Macmillan.

Laufer, B. (1997a). The lexical plight in second language reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20-34). Cambridge, England: Cambridge University Press.

Laufer, B. (1997b). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge, England: Cambridge University Press.

Laufer, B. (2000). *Task effect on instructed vocabulary learning: The hypothesis of 'involvement.'* Selected Papers from AILA '99 Tokyo. Waseda University Press, Tokyo, 47-62.

Laufer, B. (2012). Involvement load hypothesis. In P. Robinson (Ed.), *The Routledge Encyclopedia of second language acquisition* (pp. 15-27). London, England: Routledge.

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology, 3*(2), 58-76.

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task induced involvement. *Applied Linguistics, 22*(1), 1-26.

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

Laufer, B., & Paribakht, T. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning, 48*(3), 365-391.

Laufer, B., & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals, 18,* 405-411.

Leighton, J. P. & Gokiert, R. J. (2005). The cognitive effects of test item features: Informing item generation by identifying construct irrelevant variance. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Montreal, Canada.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Li, M., & Kirby, J. (2014). The effect of vocabulary breadth and depth on English reading. *Applied Linguistics, 7*, 1-25.

Linck, J. A., Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *Mental Lexicon, 3,* 349-374.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-learners. *Journal of Educational Psychology, 95,* 482-494.

Liu, Na., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal, 16*(1), 33-42.

Low, P., & Siegel, L. S. (2005). A comparison of the cognitive processes underling reading comprehension in native English and ESL speakers. *Written Language and Literacy, 8*(2), 207-231.

Macaro, E., Guo, T., Chen, H., & Tian, L. (2009). Can differential processing of L2 vocabulary inform the debate on teacher code-switching behaviour? The case of Chinese learners of English. In B. Richards, M. Daller, D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 125-146). http://dx.doi.org/10.1057/9780230242258

Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp.107-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Marschark, M., & Cornoldi, C. (1991). Imagery and verbal memory. In C. Cornoldi & M. McDaniel (Eds.), *Imagery and cognition* (pp. 133-182). New York, NY: Springer-Verlag.

Martin, C. (2011). York Assessment of Reading for Comprehension (YARC). *Educational Psychology in Practice, 27*(4), 437-439.

Martinez-Fernandez, A. (2008). Revisiting the Involvement Load Hypothesis: Awareness, type of task and type of item. Selected Proceedings of the 2007 Second Language Research Forum. http://www.lingref.com, document #1746

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 558-572.

McRae, K., Hare, M., Elman, J., & Feretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory and Cognition, 33*(1), 174-1184.

Meara, P. (1996). *The vocabulary knowledge framework*. Retrieved from http://www.lognostics.co.uk/vlibrary/meara1996c.pdf

Meara, P. (2006). Emergent properties of multilingual lexicons. *Applied Linguistics, 27*(4), 620-644.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N . Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 84-102). Cambridge, England: Cambridge University Press.

Melka- Teichroew, F. J. (1982). Receptive versus productive vocabulary: A survey. *Interlanguage Studies Bulletin, 6*, 5-33.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement (*3rd ed.), (pp. 13-103). New York, NY: Macmillan.

Messick, S. (1996). Validity and wash back in language testing. *Language Testing, 13*(3), 241-256.

Mezynski, K. (1983). Issue concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research, 53*, 253-279.

Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition, 41*(13), 197-229.

Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 47-58). Cambridge, England: Cambridge University Press.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Clevedon, UK: Multilingual Matters.

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: The link between vocabulary size and academic achievement. *Applied Linguistics Review, 4*(1), 151-172.

Murphy, K., & Green, L. (2011). Semantic priming occurs for word but not location pronunciation in the postcue task. *Cognitive Processing, 12,* 197-201.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19,* 304-330.

Nagy, W. E., Herman, P., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20,* 233-253.

Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review, 61,* 107-134.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5*(1), 12-25. Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/1983-Testing-and-teaching.pdf

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Boston, MA: Heinle & Heinle.

Nation, I. S. P. (1997). The language learning benefits of extensive reading. *The Language Teacher, 21*(5), 13-16.

Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: selection, acquisition and testing* (pp. 3-13). Amsterdam, the Netherlands: John Benjamins.

Nation, I. S. P. (2005). Vocabulary learning through extensive reading. In G. Poedjosoedarmo

(Ed.), *Innovative approaches to reading and writing* (pp 10-21). RELC Anthology Series 46. RELC, Singapore.

Nation, I. S. P. (2013a). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.

Nation, I. S. P. (2013b). Designing reading tasks to maximize vocabulary learning. *Applied Research on English Language, 3*(1), 1-8.

Nation, I. S. P., & Gu, P. Y. (2007). *Focus on vocabulary*. Sydney, Australia: National Centre for English Language Teaching and Research, Macquarie University.

Nation, P. (1993). Using dictionaries to estimate vocabulary size: Essential but rarely followed procedure. *Language Testing, 10,* 27-40.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-81.

Nation, P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton & J. Treffers-Dalle (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 33-43). Cambridge, England: Cambridge University Press.

Nation, P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal, 42*(1), 86-99.

Nation, P. (2012). Measuring vocabulary size in an uncommonly taught language. Paper presented at the International Conference on Language Proficiency Testing in the Less Commonly Taught Languages, Bangkok, Thailand

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge, England: Cambridge University Press.

Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 151-171.

Nergis, A. (2013). Exploring the factors that affect reading comprehension of EAP learners. *Journal of English for Academic Purposes, 12* (1), 1-9.

Nguyen, L., & Nation, I. S. P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal, 42*(1), 86-99.

Nitsch, K. E. (1978). *Structuring decontextualized forms of knowledge*. (Unpublished doctoral Dissertation). Vanderbilt University, Nashville, TN.

Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2014). Visual and auditory digit-span performance in native and non-native speakers. *International Journal of Bilingualism, 18,* 663-673. http://dx.doi.org/10.1177/1367006912466314

Ordóñez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology, 94,* 719 - 728. http:// dx.doi.org/10.1037/0022-0663.94.4.719

Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinehart and Winston.

Paivio, A. (1983). The empirical case for dual-coding. In J. Yuille (Ed.), *Imagery, memory and cognition: Essays in honor of Allan Paivio* (pp. 307–332). Hillsdale, NJ: Erlbaum.

Paivio, A. (1986). *Mental representation: A dual-coding approach*. New York, NY: Oxford University Press.

Paivio, A. (1995). Imagery and memory. In M. S. Gazzaniga (Ed.), *The cognitive Neurosciences* (pp. 977-986). Cambridge, MA: MIT Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement, 76,* 1-25.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology 48,* 98-107.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174-200).

Park, E. S. (2005). On three potential sources of comprehensible input for second language acquisition. *Working Papers in TESOL & Applied Linguistics, 2*(3).

Paulson, J., & Henery, J. (2002). Does the degree of reading power assessment reflect the reading process? An eye-movement examination: The Modifies Cloze Procedure demands an interruption of normal reading processes. *Journal of Adolescent & Adult Literacy, 46*(l).

Pellicer-Sánchez. A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language, 22*(1), 31-55.

Perfetti, C. A. (1985). *Reading ability*. Oxford, UK: Oxford University Press.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-83.

Phelan, C., & Wren, J., 2006. *Exploring reliability in academic assessment.* UNI Office of Academic Assessment.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1-28.

Pimsleur, P. (1967). A memory schedule. *The Modern Language Journal, 51*(2), 73-75. Retrieved from http://www.jstor.org/stable/321812

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language, 5*(2), 271-275.

Pliatsikas, C., & Marinis, T. (2013). Rule-based morphological processing in a second language*:* A behavioural investigation. *Language Studies Working Papers, 5*, 3-12.

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning*, *53*(2), 233-284.

Qian, D. (1998). *Depth of vocabulary knowledge: Assessing its role in adults' reading comprehension in English as a second language*. (Unpublished doctoral dissertation). University of Toronto, Toronto, Canada.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*(2), 282-308.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Qian, D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*(1), 28-52.

Rashidi. N., & Khosravi, N. (2010). Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics, 14*(1), 81-108.

Rastle, K., Davis, M. H., Marslen-Wilson, W., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes, 15*, 507-537.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal, 19*(2), 12-25.

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: Lawrence Erlbaum.

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

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language:*

Selection, acquisition and testing (pp. 209-227). Amsterdam, the Netherlands: Benjamins.

Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies, 7*(2), 105-125.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-89.

Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice.* Cambridge, England: Cambridge University Press.

Richardson, J. T. (2007). Measures of short-term memory: A historical review. Cortex, *43*(5), 635-650.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition, 21*(1), 589-619.

Ryan, A., & Meara, P. (1991). The case of the invisible vowels: Arabic speaker reading English words. *Reading a Foreign Language, 7,* 531-540.

Sabatini, J. P., Sawaki, Y., Shore, J. R. and Scarborough, H. S. (2010). Relationships among reading skills of adults with low literacy. *Journal of Learning Disabilities, 43*(2), 122-138.

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics, 24,* 431-51.

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System, 6*(2), 72-78.

Scarborough, D, L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Learning & Memory, 3,* 1-17.

Scarcella, R. (1997). English learners and writing: Responding to linguistic diversity. In C. B. Olson (Ed.), *Practical ideas for teaching writing as a process at the high school and college levels* (pp. 85-91). Sacramento, CA: California Department of Education.

Schellings, G., Aarnoutse, C., & Leeuwe, J. V. (2006). Third-graders' think-aloud protocols: Types of reading activities in reading an expository text. *Learning and Instruction, 16*(6), 549-569.

Schmitt, N. (1998). Tracking the incidental acquisition of second language vocabulary: A longitudinal study. *Language Learning, 48*(2), 281-317.

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge, England: Cambridge University Press.

Schmitt, N. (2007). Current perspectives on vocabulary teaching and learning. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching: Part II* (pp. 827-841). New York, NY: Springer.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition, 19,* 17-36.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.

Schmitt, N., Xiangying, J., & William, G. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26-43.

Schmidt, R. (1994). Deconstructing consciousness in search for useful definitions for applied linguistics. *AILA Review, 11,* 11-26.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45,* 21-50.

Shimron, J. (1999). The role of vowel signs in Hebrew: Beyond word recognition. *Reading and Writing: An Interdisciplinary Journal, 11,* 301-319.

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

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition, 11,* 245-260.

Staehr, L. (2009). Vocabulary knowledge and advanced listening comprehension in English as a Foreign Language. *Studies in Second Language Acquisition, 31,* 1-31.

Stanovich, E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-407.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stothard, S. E., Snowling, M. J., Clarke, P. J., Barmby, P., & Hulme, C. (2010). York assessment of reading and comprehension (2nd ed.). London: GL Assessment.

Tan, D., Pandian, A., & Jaganathan, P. (2016). Lexical testing and the reliability of the modified Vocabulary Knowledge Scale. *Advances in Language and Literary Studies, 7,* (5), 91-96.

Tang, C., & Treffers-Daller, J. (2016). Assessing incidental vocabulary learning by Chinese EFL learners: A test of the Involvement Load Hypothesis. In G. Yu & Y. Yin (Eds.), *Assessing Chinese learners of English: Language constructs, consequences and conundrums* (pp. 121-148). London, UK: Palgrave.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10*(4), 381-398.

Thorn, A. S. C., & Gathercole, S. E. (2001). Language differences in verbal short-term memory do not exclusively originate in the process of subvocal rehearsal. *Psychonomic Bulletin & Review, 8,* 357-364.

Tsai, Y. R., Ernst, C., & Talley, P. C. (2010). L1 and L2 strategy use in reading comprehension of Chinese EFL readers. *Reading Psychology, 31*(1), pp.1-29.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the Independent Components Hypothesis. *Journal of Learning Disabilities, 45*(5), 453-466.

Ulijn, J. M., & Kempen, G. A. M. (1976). The role of the first language in second language reading comprehension: Some experimental evidence. In G. Nickel (Ed.), Proceedings of the 4th International Congress of Applied Linguistics, Vol. 1. Stuttgart, Germany: Hochschulverlag.

Vela, V. (2014). Second language vocabulary acquisition from a linguistic point of view. *The Journal of Teaching English for Specific and Academic Purposes, 2,* 293-303.

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology, 22*(3), 407-423.

Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. In *Kiyo: The occasional papers of Notre Dame Seishin University*, Japan.

Waring, R., & Nation, I. S. P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English Speaking World, 4*, 97-110.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*, 130-163.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(1), 33-52.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York, NY: The Psychological Corporation.

Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: The Psychological Corporation,

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition*. San Antonio, TX: Pearson.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53,* 13-40.

West, M. (1953). *A general service list of English words*. London, England: Longman.

West, R. F., & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science, 2*(5), 325-329.

Wolter, B. (2005). *V_Links: A new approach to assessing depth of word knowledge*. (Unpublished PhD dissertation). University of Wales, Swansea.

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L2 lexical/conceptual knowledge. *Applied Linguistics, 27*(4), 741-747.

Yaqubi, B., Rayati, R. A., & Allemzade Gorgi, N. (2010). The involvement load hypothesis and vocabulary learning: The effect of task types and involvement index on L2 vocabulary acquisition. *The Journal of Teaching Language Skills, 2*(1), 145-163.

Zareva, A, Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition, 27,* 567-595.

Zechmeister, E. B., Chronis, A. M., Cull, W. L., D'Anna, C. A., & Healy, N. A. (1995). Growth of a functionally important lexicon. *Journal of Reading Behavior, 27*(2), 201-212.

Zhang, D., & Yang, X. (2016). Chinese L2 learners' depth of vocabulary knowledge and its role in reading comprehension. *Foreign Language Annals, 49*(4), 699-715.

Zhong, H. F. (2016). The relationship between receptive and productive vocabulary knowledge: A perspective from vocabulary use in sentence writing. *The Language Learning Journal*, 1-17. http://dx.doi.org/10.1080/09571736.2015.1127403

Appendixes

Appendix 1: Participation Consent Forms for L1 Arabic learners of English

56

Consent Form 1

I have read and had explained to me by Reham Ibrahim Alkhudairy the Information Sheet relating to this project.

I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

1. Name: Noura Abdulrhman AL-Shedokhi

2. Native language(s): Arabic

3. Gender: (a) female (b) male (tick one)

4. Dept/ year: English / seven

5. Date of birth: 22/10/1996

6. Age at which I first started learning English (for non-native English speakers): 10

7. Number of years you have attended English classes (for non-native English speakers): 11 years

Date: 23/02/2015

Sign:

Notice: This information will be kept strictly confidential, and is only to help me with the analysis of the results. Once the results have been analysed, all personal information held by me about you will be deleted.

Appendix 1 (Cont.)

(37)

Consent Form

1

I have read and had explained to me by Reham Ibrahim Alkhudairy the Information Sheet relating to this project.

I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

1. Name: AlAnoud AlNoudali

2. Native language(s): Arabic

3. Gender: ~~(a)~~ female (b) male (tick one)

4. Dept/ year: English Level: 6

5. Date of birth: 6/3/1994 / 20

6. Age at which I first started learning English (for non-native English speakers): Sixth grade. 12

7. Number of years you have attended English classes (for non-native English speakers): 9 years

Date: 28/02/2015

Sign: [signature]

Notice: This information will be kept strictly confidential, and is only to help me with the analysis of the results. Once the results have been analysed, all personal information held by me about you will be deleted.

Appendix 2a: Vocabulary Size Test

Circle the letter a-d with the closest meaning to the key word in the question.

1. COMPOUND: They made a new **compound**.
   a. agreement
   b. thing made of two or more parts
   c. group of people forming a business
   d. guess based on past experience

2. LATTER: I agree with the **latter**.
   a. man from the church
   b. reason given
   c. last one
   d. answer

3. CANDID: Please be **candid**.
   a. be careful
   b. show sympathy
   c. show fairness to both sides
   d. say what you really think

4. TUMMY: Look at my **tummy**.
   a. cloth to cover the head
   b. stomach
   c. small furry animal
   d. thumb

5. QUIZ: We made a **quiz**.
   a. thing to hold arrows
   b. serious mistake
   c. set of questions
   d. box for birds to make nests in

6. INPUT: We need more **input**.
   a. information, power, etc. put into something
   b. workers
   c. artificial filling for a hole in wood
   d. money

7. CRAB: Do you like **crabs**?
   a. sea creatures that walk sideways
   b. very thin small cakes
   c. tight, hard collars
   d. large black insects that sing at night

8. VOCABULARY: You will need more **vocabulary**.
   a. words
   b. skill
   c. money
   d. guns

9. REMEDY: We found a good **remedy**.
   a. way to fix a problem
   b. place to eat in public
   c. way to prepare food
   d. rule about numbers

10. ALLEGE: They **alleged** it.
    a. claimed it without proof
    b. stole the ideas for it from someone else
    c. provided facts to prove it
    d. argued against the facts that supported it

1. PRO: He's a **pro**.
   a. someone who is employed to find out important secrets
   b. a stupid person
   c. someone who writes for a newspaper
   d. someone who is paid for playing sport etc

2. MICROPONE: Please use the **microphone**.
   a. machine for making food hot
   b. machine that makes sounds louder
   c. machine that makes things look bigger
   d. small telephone that can be carried around

3. CIRCLE: Make a **circle**.
   a. rough picture
   b. space with nothing in it
   c. round shape
   d. large hole

4. PUB: They went to the **pub**.
   a. place where people drink and talk
   b. place that looks after money
   c. large building with many shops
   d. building for swimming

5. NIL: His mark for that question was **nil**.
   a. very bad
   b. nothing
   c. very good
   d. in the middle

6. PATIENCE: He has no **patience**.
   a. will not wait happily
   b. has no free time
   c. has no faith
   d. does not know what is fair

7. DRAWER: The **drawer** was empty.
   a. sliding box
   b. place where cars are kept
   c. cupboard to keep things cold
   d. animal house

8. UPSET: I am **upset**.
   a. tired
   b. famous
   c. rich
   d. unhappy

9. STONE: He sat on a **stone**.
   a. hard thing
   b. kind of chair
   c. soft thing on the floor
   d. part of a tree

10. MAINTAIN: Can they **maintain** it?
    a. keep it as it is
    b. make it larger
    c. get a better one than it
    d. get it

Appendix 2a (Cont.)

1. ERRATIC: He was **erratic**.
   a. without fault
   b. very bad
   c. very polite
   d. unsteady

2. PALETTE: He lost his **palette**.
   a. basket for carrying fish
   b. wish to eat food
   c. young female companion
   d. artist's board for mixing paints

3. NULL: His influence was **null**.
   a. had good results
   b. was unhelpful
   c. had no effect
   d. was long-lasting

4. KINDERGARTEN: This is a good **kindergarten**.
   a. activity that allows you to forget your worries
   b. place of learning for children too young for school
   c. strong, deep bag carried on the back
   d. place where you may borrow books

5. ECLIPSE: There was an **eclipse**.
   a. a strong wind
   b. a loud noise of something hitting the water
   c. The killing of a large number of people
   d. The sun hidden by a planet

6. MARROW: This is the **marrow**.
   a. symbol that brings good luck to a team
   b. Soft centre of a bone
   c. control for guiding a plane
   d. increase in salary

7. LOCUST: There were hundreds of **locusts**.
   a. insects with wings
   b. unpaid helpers
   c. people who do not eat meat
   d. brightly coloured wild flowers

8. AUTHENTIC: It is **authentic**.
   a. real
   b. very noisy
   c. Old
   d. Like a desert

9. CABARET: We saw the **cabaret**.
   a. painting covering a whole wall
   b. song and dance performance
   c. small crawling insect
   d. person who is half fish, half woman

10. MUMBLE: He started to **mumble**.
    a. think deeply
    b. shake uncontrollably
    c. stay further behind the others
    d. speak in an unclear way

1. VEER: The car **veered**.
   a. went suddenly in another direction
   b. moved shakily
   c. made a very loud noise
   d. slid sideways without the wheels turning

2. MALIGN: His **malign** influence is still felt.
   a. evil
   b. good
   c. very important
   d. secret

3. CAVALIER: He treated her in a **cavalier** manner.
   a. without care
   b. politely
   c. awkwardly
   d. as a brother would

4. STRANGLE: He **strangled** her.
   a. killed her by pressing her throat
   b. gave her all the things she wanted
   c. took her away by force
   d. admired her greatly

5. THESIS: She has completed her **thesis**.
   a. long written report of study carried out for a university degree
   b. talk given by a judge at the end of a trial
   c. first year of employment after becoming a teacher
   d. extended course of hospital treatment

6. THRESHOLD: They raised the **threshold**.
   a. flag
   b. point or line where something changes
   c. roof inside a building
   d. cost of borrowing money

7. ACCESSORY: They gave us some **accessories**.
   a. papers allowing us to enter a country
   b. official orders
   c. ideas to choose between
   d. extra pieces

8. BUTLER: They have a **butler**.
   a. man servant
   b. machine for cutting up trees
   c. private teacher
   d. cool dark room under the house

9. PREMIER: The **premier** spoke for an hour.
   a. person who works in a law court
   b. university teacher
   c. adventurer
   d. head of the government

10. DEVIOUS: Your plans are **devious**.
    a. tricky
    b. well-developed
    c. not well thought out
    d. more expensive than necessary

Appendix 2a (Cont.)

1. SEE: They **saw** it.
   a. cut
   b. waited for
   c. looked at
   d. started

2. TIME: They have a lot of **time**.
   a. money
   b. food
   c. hours
   d. friends

3. PERIOD: It was a difficult **period**.
   a. question
   b. time
   c. thing to do
   d. book

4. FIGURE: Is this the right **figure**?
   a. answer
   b. place
   c. time
   d. number

5. POOR: We are **poor**.
   a. have no money
   b. feel happy
   c. are very interested
   d. do not like to work hard

6. DRIVE: He **drives** fast.
   a. swims
   b. learns
   c. throws balls
   d. uses a car

7. JUMP: She tried to **jump**.
   a. lie on top of the water
   b. get off the ground suddenly
   c. stop the car at the edge of the road
   d. move very fast

8. SHOE: Where is your **shoe**?
   a. the person who looks after you
   b. the thing you keep your money in
   c. the thing you use for writing
   d. the thing you wear on your foot

9. STANDARD: Her **standards** are very high.
   a. the bits at the back under her shoes
   b. the marks she gets in school
   c. the money she asks for
   d. the levels she reaches in everything

10. BASIS: This was used as the **basis**.
    a. answer
    b. place to take a rest
    c. next step
    d. main part

1. BACTERIUM: They didn't find a single **bacterium**.
   a. small living thing causing disease
   b. plant with red or orange flowers
   c. animal that carries water on its back
   d. thing that has been stolen and sold to a shop

2. FRACTURE: They found a **fracture**.
   a. break
   b. small piece
   c. short coat
   d. rare jewel

3. PEEL: Shall I **peel** it?
   a. let it sit in water for a long time
   b. take the skin off it
   c. make it white
   d. cut it into thin pieces

4. MINIATURE: It is a **miniature**.
   a. a very small thing of its kind
   b. an instrument to look at small objects
   c. a very small living creature
   d. a small line to join letters in handwriting

5. CUBE: I need one more **cube**.
   a. sharp thing used for joining things
   b. solid square block
   c. tall cup with no saucer
   d. piece of stiff paper folded in half

6. COMPOST: We need some **compost**.
   a. strong support
   b. help to feel better
   c. hard stuff made of stones and sand stuck together
   d. rotted plant material

7. HAUNT: The house is **haunted**.
   a. full of ornaments
   b. rented
   c. empty
   d. full of ghosts

8. NUN: We saw a **nun**.
   a. long thin creature that lives in the earth
   b. terrible accident
   c. woman following a strict religious life
   d. unexplained bright light in the sky

9. WEEP: He **wept**.
   a. finished his course
   b. cried
   c. died
   d. worried

10. DEFICIT: The company had a large **deficit**.
    a. spent a lot more money than it earned
    b. went down a lot in value
    c. had a plan for its spending that used a lot of money
    d. had a lot of money in the bank

Appendix 2a (Cont.)

1. OLIVE: We bought **olives**.
   a. oily fruit
   b. scented pink or red flowers
   c. men's clothes for swimming
   d. tools for digging up weeds

2. QUILT: They made a **quilt**.
   a. statement about who should get their property when they die
   b. firm agreement
   c. thick warm cover for a bed
   d. feather pen

3. STEALTH: They did it by **stealth**.
   a. spending a large amount of money
   b. hurting someone so much that they agreed to their demands
   c. moving secretly with extreme care and quietness
   d. taking no notice of problems they met

4. SHUDDER: The boy **shuddered**.
   a. spoke with a low voice
   b. almost fell
   c. shook
   d. called out loudly

5. BRISTLE: The **bristles** are too hard.
   a. questions
   b. short stiff hairs
   c. folding beds
   d. bottoms of the shoes

6. BLOC: They have joined this **bloc**.
   a. musical group
   b. band of thieves
   c. small group of soldiers who are sent ahead of others
   d. group of countries sharing a purpose

7. DEMOGRAPHY: This book is about **demography**.
   a. the study of patterns of land use
   b. the study of the use of pictures to show facts about numbers
   c. the study of the movement of water
   d. the study of population

8. GIMMICK: That's a good **gimmick**.
   a. thing for standing on to work high above the ground
   b. small thing with pockets to hold money
   c. attention-getting action or thing
   d. clever plan or trick

9. AZALEA: This **azalea** is very pretty.
   a. small tree with many flowers growing in groups
   b. light material made from natural threads
   c. long piece of material worn by women in India
   d. sea shell shaped like a fan

10. YOGHURT: This **yoghurt** is disgusting.
    a. grey mud found at the bottom of rivers
    b. unhealthy, open sore
    c. thick, soured milk, often with sugar and flavouring
    d. large purple fruit with soft flesh

1. LONESOME: He felt **lonesome**.
   a. ungrateful
   b. very tired
   c. lonely full of energy
   d. full of energy

2. ROVE: He couldn't stop **roving**.
   a. getting drunk
   b. travelling around
   c. making a musical sound through closed lips
   d. working hard

3. DASH: They **dashed** over it.
   a. moved quickly
   b. moved slowly
   c. fought
   d. looked quickly

4. PAVE: It was **paved**.
   a. prevented from going through
   b. divided
   c. given gold edges
   d. covered with a hard surface

5. STRAP: He broke the **strap**.
   a. promise
   b. top cover
   c. shallow dish for food
   d. strip of material for holding things together

6. DINOSAUR: The children were pretending to be **dinosaurs**.
   a. robbers who work at sea
   b. very small creatures with human form but with wings
   c. large creatures with wings that breathe fire
   d. animals that lived a long time ago

7. SCRUB: He is **scrubbing** it.
   a. cutting shallow lines into it
   b. repairing it
   c. rubbing it hard to clean it
   d. drawing simple pictures of it

8. JUG: He was holding a **jug**.
   a. A container for pouring liquids
   b. an informal discussion
   c. A soft cap
   d. A weapon that explodes

9. RESTORE: It has been **restored**.
   a. said again
   b. given to a different person
   c. given a lower price
   d. made like new again

10. SOLDIER: He is a **soldier**.
    a. person in a business
    b. student
    c. person who uses metal
    d. person in the army

Appendix 2b: Vocabulary Size Test

Circle the letter a-d with the closest meaning to the key word in the question.

1. YOGHURT: This **yoghurt** is disgusting.
   a. grey mud found at the bottom of rivers
   b. unhealthy, open sore
   c. thick, soured milk, often with sugar and flavouring
   d. large purple fruit with soft flesh

2. AZALEA: This **azalea** is very pretty.
   a. small tree with many flowers growing in groups
   b. light material made from natural threads
   c. long piece of material worn by women in India
   d. sea shell shaped like a fan

3. GIMMICK: That's a good **gimmick**.
   a. thing for standing on to work high above the ground
   b. small thing with pockets to hold money
   c. attention-getting action or thing
   d. d. clever plan or trick

4. DEMOGRAPHY: This book is about **demography**.
   a. the study of patterns of land use
   b. the study of the use of pictures to show facts about numbers
   c. the study of the movement of water
   d. the study of population

5. BLOC: They have joined this **bloc**.
   a. musical group
   b. band of thieves
   c. small group of soldiers who are sent ahead of others
   d. group of countries sharing a purpose

6. BRISTLE: The **bristles** are too hard.
   a. questions
   b. short stiff hairs
   c. folding beds
   d. bottoms of the shoes

7. SHUDDER: The boy **shuddered**.
   a. spoke with a low voice
   b. almost fell
   c. shook
   d. called out loudly

8. STEALTH: They did it by **stealth**.
   a. spending a large amount of money
   b. hurting someone so much that they agreed to their demands
   c. moving secretly with extreme care and quietness
   d. taking no notice of problems they met

9. QUILT: They made a **quilt**.
   a. statement about who should get their property when they die
   b. firm agreement
   c. thick warm cover for a bed
   d. feather pen

10. OLIVE: We bought **olives**.
    a. oily fruit
    b. scented pink or red flowers
    c. men's clothes for swimming
    d. tools for digging up weeds

1. SOLDIER: He is a **soldier**.
   a. person in a business
   b. student
   c. person who uses metal
   d. person in the army

2. RESTORE: It has been **restored**.
   a. said again
   b. given to a different person
   c. given a lower price
   d. made like new again

3. JUG: He was holding a **jug**.
   a. A container for pouring liquids
   b. an informal discussion
   c. A soft cap
   d. A weapon that explodes

4. SCRUB: He is **scrubbing** it.
   a. cutting shallow lines into it
   b. repairing it
   c. rubbing it hard to clean it
   d. drawing simple pictures of it

5. DINOSAUR: The children were pretending to be **dinosaurs**.
   a. robbers who work at sea
   b. very small creatures with human form but with wings
   c. large creatures with wings that breathe fire
   d. animals that lived a long time ago

6. STRAP: He broke the **strap**.
   a. promise
   b. top cover
   c. shallow dish for food
   d. strip of material for holding things together

7. PAVE: It was **paved**.
   a. prevented from going through
   b. divided
   c. given gold edges
   d. covered with a hard surface

8. DASH: They **dashed** over it.
   a. moved quickly
   b. moved slowly
   c. fought
   d. looked quickly

9. ROVE: He couldn't stop **roving**.
   a. getting drunk
   b. travelling around
   c. making a musical sound through closed lips
   d. working hard

10. LONESOME: He felt **lonesome**.
    a. ungrateful
    b. very tired
    c. lonely
    d. full of energy

Appendix 2b (Cont.)

1. BASIS: This was used as the **basis**.
   a. answer
   b. place to take a rest
   c. next step
   d. main part

2. STANDARD: Her **standards** are very high.
   a. the bits at the back under her shoes
   b. the marks she gets in school
   c. the money she asks for
   d. the levels she reaches in everything

3. SHOE: Where is your **shoe**?
   a. the person who looks after you
   b. the thing you keep your money in
   c. the thing you use for writing
   d. the thing you wear on your foot

4. JUMP: She tried to **jump**.
   a. lie on top of the water
   b. get off the ground suddenly
   c. stop the car at the edge of the road
   d. d. move very fast

5. DRIVE: He **drives** fast.
   a. swims
   b. learns
   c. throws balls
   d. uses a car

6. POOR: We are **poor**.
   a. have no money
   b. feel happy
   c. are very interested
   d. do not like to work hard

7. FIGURE: Is this the right **figure**?
   a. answer
   b. place
   c. time
   d. number

8. PERIOD: It was a difficult **period**.
   a. question
   b. time
   c. thing to do
   d. book

9. TIME: They have a lot of **time**.
   a. money
   b. food
   c. hours
   d. friends

10. SEE: They **saw** it.
    a. cut
    b. waited for
    c. looked at
    d. started

1. DEFICIT: The company had a large **deficit**.
   a. spent a lot more money than it earned
   b. went down a lot in value
   c. had a plan for its spending that used a lot of money
   d. had a lot of money in the bank

2. WEEP: He **wept**.
   a. finished his course
   b. cried
   c. died
   d. worried

3. NUN: We saw a **nun**.
   a. long thin creature that lives in the earth
   b. terrible accident
   c. woman following a strict religious life
   d. unexplained bright light in the sky

4. HAUNT: The house is **haunted**.
   a. full of ornaments
   b. rented
   c. empty
   d. full of ghosts

5. COMPOST: We need some **compost**.
   a. strong support
   b. help to feel better
   c. hard stuff made of stones and sand stuck together
   d. rotted plant material

6. CUBE: I need one more **cube**.
   a. sharp thing used for joining things
   b. solid square block
   c. tall cup with no saucer
   d. piece of stiff paper folded in half

7. MINIATURE: It is a **miniature**.
   a. a very small thing of its kind
   b. an instrument to look at small objects
   c. a very small living creature
   d. a small line to join letters in handwriting

8. PEEL: Shall I **peel** it?
   a. let it sit in water for a long time
   b. take the skin off it
   c. make it white
   d. cut it into thin pieces

9. FRACTURE: They found a **fracture**.
   a. break
   b. small piece
   c. short coat
   d. rare jewel

10. BACTERIUM: They didn't find a single **bacterium**.
    a. small living thing causing disease
    b. plant with red or orange flowers
    c. animal that carries water on its back
    d. thing that has been stolen and sold to a shop

Appendix 2b (Cont.)

1. MUMBLE: He started to **mumble**.
   a. think deeply
   b. shake uncontrollably
   c. stay further behind the others
   d. speak in an unclear way

2. CABARET: We saw the **cabaret**.
   a. painting covering a whole wall
   b. song and dance performance
   c. small crawling insect
   d. person who is half fish, half woman

3. AUTHENTIC: It is **authentic**.
   a. real
   b. very noisy
   c. Old
   d. Like a desert

4. LOCUST: There were hundreds of **locusts**.
   a. insects with wings
   b. unpaid helpers
   c. people who do not eat meat
      brightly
   d. coloured wild flowers

5. MARROW: This is the **marrow**.
   a. symbol that brings good luck to a team
   b. Soft centre of a bone
   c. control for guiding a plane
   d. increase in salary

6. ECLIPSE: There was an **eclipse**.
   a. a strong wind
   b. a loud noise of something hitting the water
   c. The killing of a large number of people
   d. d.The sun hidden by a planet

7. KINDERGARTEN: This is a good **kindergarten**.
   a. activity that allows you to forget your worries
   b. place of learning for children too young for school
   c. strong, deep bag carried on the back
   d. d. place where you may borrow books

8. NULL: His influence was **null**.
   a. had good results
   b. was unhelpful
   c. had no effect
   d. was long-lasting

9. PALETTE: He lost his **palette**.
   a. basket for carrying fish
   b. wish to eat food
   c. young female companion
   d. artist's board for mixing paints

10. ERRATIC: He was **erratic**.
    a. without fault
    b. very bad
    c. very polite
    d. unsteady

1. DEVIOUS: Your plans are **devious**.
   a. tricky
   b. well-developed
   c. not well thought out
   d. more expensive than necessary

2. PREMIER: The **premier** spoke for an hour.
   a. person who works in a law court
   b. university teacher
   c. adventurer
   d. head of the government

3. BUTLER: They have a **butler**.
   a. man servant
   b. machine for cutting up trees
   c. private teacher
   d. cool dark room under the house

4. ACCESSORY: They gave us some **accessories**.
   a. papers allowing us to enter a country
   b. official orders
   c. ideas to choose between
   d. extra pieces

5. THRESHOLD: They raised the **threshold**.
   a. flag
   b. point or line where something changes
   c. roof inside a building
   d. cost of borrowing money

6. THESIS: She has completed her **thesis**.
   a. long written report of study carried out for a university degree
   b. talk given by a judge at the end of a trial
   c. first year of employment after becoming a teacher
   d. extended course of hospital treatment

7. STRANGLE: He **strangled** her.
   a. killed her by pressing her throat
   b. gave her all the things she wanted
   c. took her away by force
   d. admired her greatly

8. CAVALIER: He treated her in a **cavalier** manner.
   a. without care
   b. politely
   c. awkwardly
   d. as a brother would

9. MALIGN: His **malign** influence is still felt.
   a. evil
   b. good
   c. very important
   d. secret

10. VEER: The car **veered**.
    a. went suddenly in another direction
    b. moved shakily
    c. made a very loud noise
    d. slid sideways without the wheels turning

Appendix 2b (Cont.)

1. ALLEGE: They **alleged** it.
   a. claimed it without proof
   b. stole the ideas for it from someone else
   c. provided facts to prove it
   d. argued against the facts that supported it

2. REMEDY: We found a good **remedy**.
   a. way to fix a problem
   b. place to eat in public
   c. way to prepare food
   d. rule about numbers

3. VOCABULARY: You will need more **vocabulary**.
   a. words
   b. skill
   c. money
   d. guns

4. CRAB: Do you like **crabs**?
   a. sea creatures that walk sideways
   b. very thin small cakes
   c. tight, hard collars
   d. large black insects that sing at night

5. INPUT: We need more **input**.
   a. information, power, etc. put into something
   b. workers
   c. artificial filling for a hole in wood
   d. money

6. QUIZ: We made a **quiz**.
   a. thing to hold arrows
   b. serious mistake
   c. set of questions
   d. box for birds to make nests in

7. TUMMY: Look at my **tummy**.
   a. cloth to cover the head
   b. stomach
   c. small furry animal
   d. thumb

8. CANDID: Please be **candid**.
   a. be careful
   b. show sympathy
   c. show fairness to both sides
   d. say what you really think

9. LATTER: I agree with the **latter**.
   a. man from the church
   b. reason given
   c. last one
   d. answer

10. COMPOUND: They made a new **compound**.
    a. agreement
    b. thing made of two or more parts
    c. group of people forming a business
    d. guess based on past experience

1. MAINTAIN: Can they **maintain** it?
   a. keep it as it is
   b. make it larger
   c. get a better one than it
   d. get it

2. STONE: He sat on a **stone**.
   a. hard thing
   b. kind of chair
   c. soft thing on the floor
   d. part of a tree

3. UPSET: I am **upset**.
   a. tired
   b. famous
   c. rich
   d. unhappy

4. DRAWER: The **drawer** was empty.
   a. sliding box
   b. place where cars are kept
   c. cupboard to keep things cold
   d. animal house

5. PATIENCE: He has no **patience**.
   a. will not wait happily
   b. has no free time
   c. has no faith
   d. does not know what is fair

6. NIL: His mark for that question was **nil**.
   a. very bad
   b. nothing
   c. very good
   d. in the middle

7. PUB: They went to the **pub**.
   a. place where people drink and talk
   b. place that looks after money
   c. large building with many shops
   d. building for swimming

8. CIRCLE: Make a **circle**.
   a. rough picture
   b. space with nothing in it
   c. round shape
   d. large hole

9. MICROPONE: Please use the **microphone**.
   a. machine for making food hot
   b. machine that makes sounds louder
   c. machine that makes things look bigger
   d. small telephone that can be carried around

10. PRO: He's a **pro**.
    a. someone who is employed to find out important secrets
    b. a stupid person
    c. someone who writes for a newspaper
    d. someone who is paid for playing sport etc

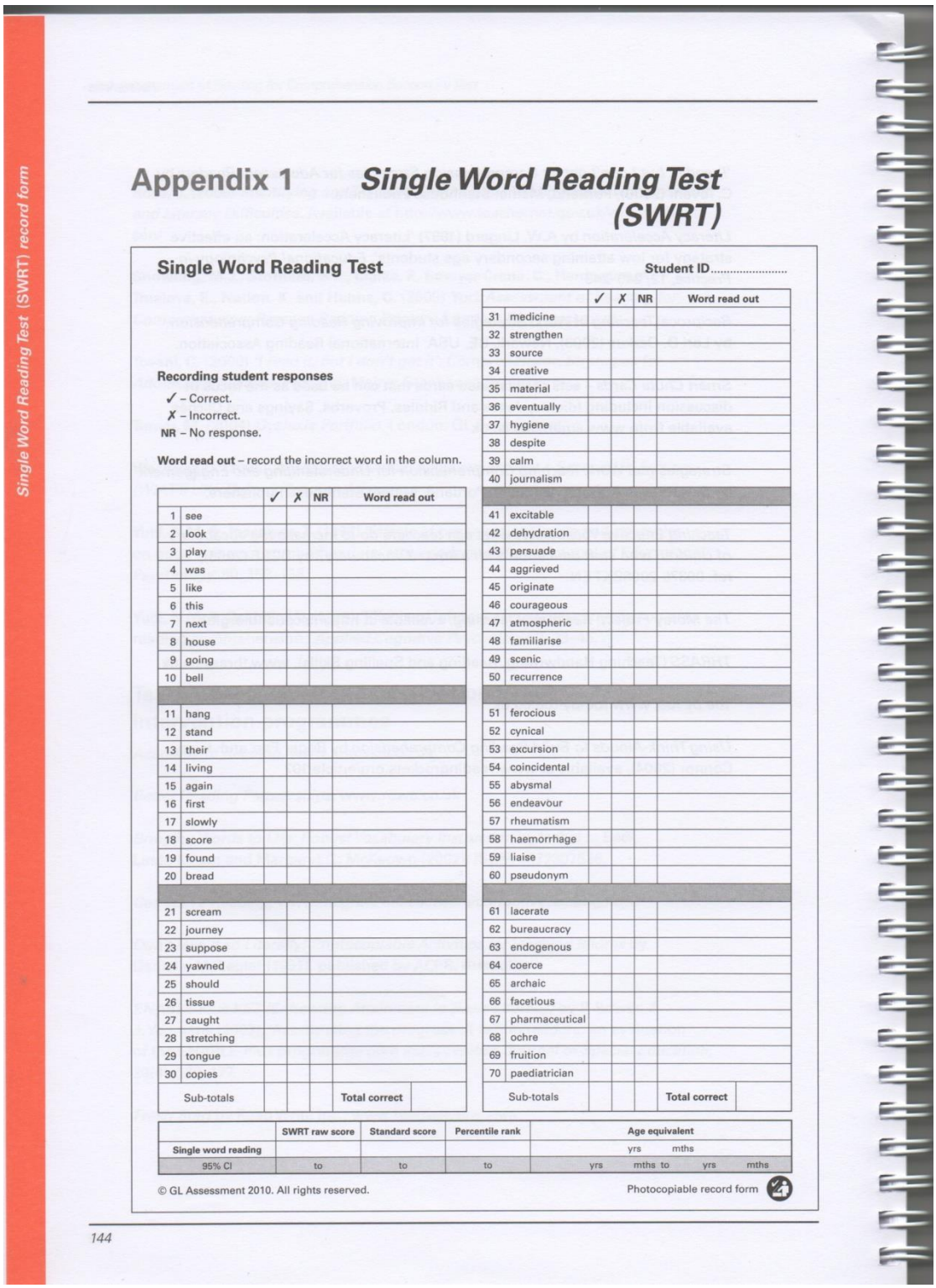
Appendix 3: Vocabulary Knowledge Scale (Chapter 3)

Name:.....

A= I know what this word/ phrase means and I can use it in a sentence.
B= I know what this word/ phrase means, but I am not sure how to use it.
C= I've seen this word/ phrase before, but I don't know what it means.
D= I've never seen this word/ phrase before.

	A	B	C	D
beauty				
pocoko				
tissue				
cenedies				
weather				
bicycle				
sataca				
minute				
toroko				
candle				

Appendix 4a: The York Assessment of Reading for Comprehension (SWRT)



Single Word Reading Test (SWRT) record form

Appendix 1 *Single Word Reading Test (SWRT)*

Single Word Reading Test

Student ID.....

Recording student responses

✓ – Correct.
✗ – Incorrect.
NR – No response.

Word read out – record the incorrect word in the column.

		✓	✗	NR	Word read out
1	see				
2	look				
3	play				
4	was				
5	like				
6	this				
7	next				
8	house				
9	going				
10	bell				
11	hang				
12	stand				
13	their				
14	living				
15	again				
16	first				
17	slowly				
18	score				
19	found				
20	bread				
21	scream				
22	journey				
23	suppose				
24	yawned				
25	should				
26	tissue				
27	caught				
28	stretching				
29	tongue				
30	copies				
	Sub-totals				Total correct

		✓	✗	NR	Word read out
31	medicine				
32	strengthen				
33	source				
34	creative				
35	material				
36	eventually				
37	hygiene				
38	despite				
39	calm				
40	journalism				
41	excitable				
42	dehydration				
43	persuade				
44	aggrieved				
45	originate				
46	courageous				
47	atmospheric				
48	familiarise				
49	scenic				
50	recurrence				
51	ferocious				
52	cynical				
53	excursion				
54	coincidental				
55	abysmal				
56	endeavour				
57	rheumatism				
58	haemorrhage				
59	liaise				
60	pseudonym				
61	lacerate				
62	bureaucracy				
63	endogenous				
64	coerce				
65	archaic				
66	facetious				
67	pharmaceutical				
68	ochre				
69	fruition				
70	paediatrician				
	Sub-totals				Total correct

	SWRT raw score	Standard score	Percentile rank	Age equivalent
Single word reading				yrs mths
95% CI	to	to	to	yrs mths to yrs mths

© GL Assessment 2010. All rights reserved.

Photocopiable record form

Appendix 4b: Age equivalent scores

Age equivalent scores

Table 2.4 gives the conversion of raw scores to reading age equivalent scores. Reading ages are the ages at which a given raw score is average.

Table 2.4: Conversion of SWRT *raw scores to age equivalent scores*

SWRT raw score	Age equivalent (years:months)
0–10	<6:00
11–13	6:00
14–17	6:03
18–19	6:06
20–22	6:09
23	7:00
24–25	7:03
26–27	7:06
28–29	7:09
30–31	8:00
32–33	8:03
34	8:06
35–36	8:09
37	9:00
38–39	9:03
40	9:06
41	9:09
42	10:00
43	10:03
44	10:06
45	10:09
46	11:00
47	11:03
48	11:09
49	12:00
50	12:06
51	12:09
52	13:03
53	13:09
54	14:00
55	14:06
56	14:09
57	15:03
58	15:09
59–70	>16:00

Table 2.4: Conversion of SWRT raw scores to age equivalent scores

15

Appendix 4c: The York Assessment of Reading for Comprehension (Passage A)

Honey for You, Honey for Me

In Southern Africa there is a bird called the Honey Guide. It is a small bird with a long pink nose. Its favourite food is honey. From a distance, the Honey Guide looks colourless and brown, but up close you can see a splash of light yellow on its white chest feathers. The bird looks a little as if it has just enjoyed a meal of golden honey, and been none too careful about its table manners! However, the Honey Guide gets its name not just from the colour of its chest; it is very well adapted to feeding on the contents of bees' nests. It eats the bees' honey, their eggs and even the bee's wax. In fact, they are one of the only birds that can eat wax. The Honey Guide is what you might call a bee specialist.

It does, however, have one major problem: the bees' **toroko**. The Honey Guide is not a big bird, and a bee **toroko** can be very dangerous, or even kill it so the bird has to find a way to enter the bees' nest and avoid a bad **toroko**.

The Honey Guide has developed a very beautiful solution to the problem. It uses **cenedies**. The Honey Guide searches around its county in the African green lands until it finds a likely-looking bees' nest. When it has found one, it flies off to find some helpers.

The bird attracts the attention of **cenedies** with a non-stop song with short sounds. Once the bird is sure that it has their full attention, it begins to fly off towards the bees' nest, stopping along the way to check that the **cenedies** are following. As the bird flies, it fans its tail out wide, to make it easier to see. When the bird and its helpers reach the bees' nest, the bird calls again to let them know that they have arrived.

Cenedies are well fitted to deal with a nest of angry bees. Using smoke, they drug the bees so they will not be hurt, then break open the nest with a knife to get at the honey pieces. The **cenedies** take as much of the honey as they want, but always leave some for the Honey Guide. In this way the bird has its meal without risk of a harmful **toroko**, and the **cenedies** get help finding hard-to-find bees' nests.

There is a saying among some families that, if the **cenedies** are hungry and do not leave a gift of honey for the Honey Guide, the bird will get angry. If that happens, they say, the next time the bird will guide the **cenedies** not to a bees' nest, but to a lion, an elephant or a harmful snake. It does not pay to be thankless for help, even when your helper is a little bird!

Appendix 4d: Reading Comprehension Questions of Passage A

Question	Student's response	Score
1. Suggest three reasons why the bird is called a Honey Guide?		1 0 NR
2. What three colours are its feathers?		1 0 NR
3. Why does the Honey Guide look colourless from a distance?		1 0 NR
4. Why are the bees so dangerous to the Honey Guide?		1 0 NR

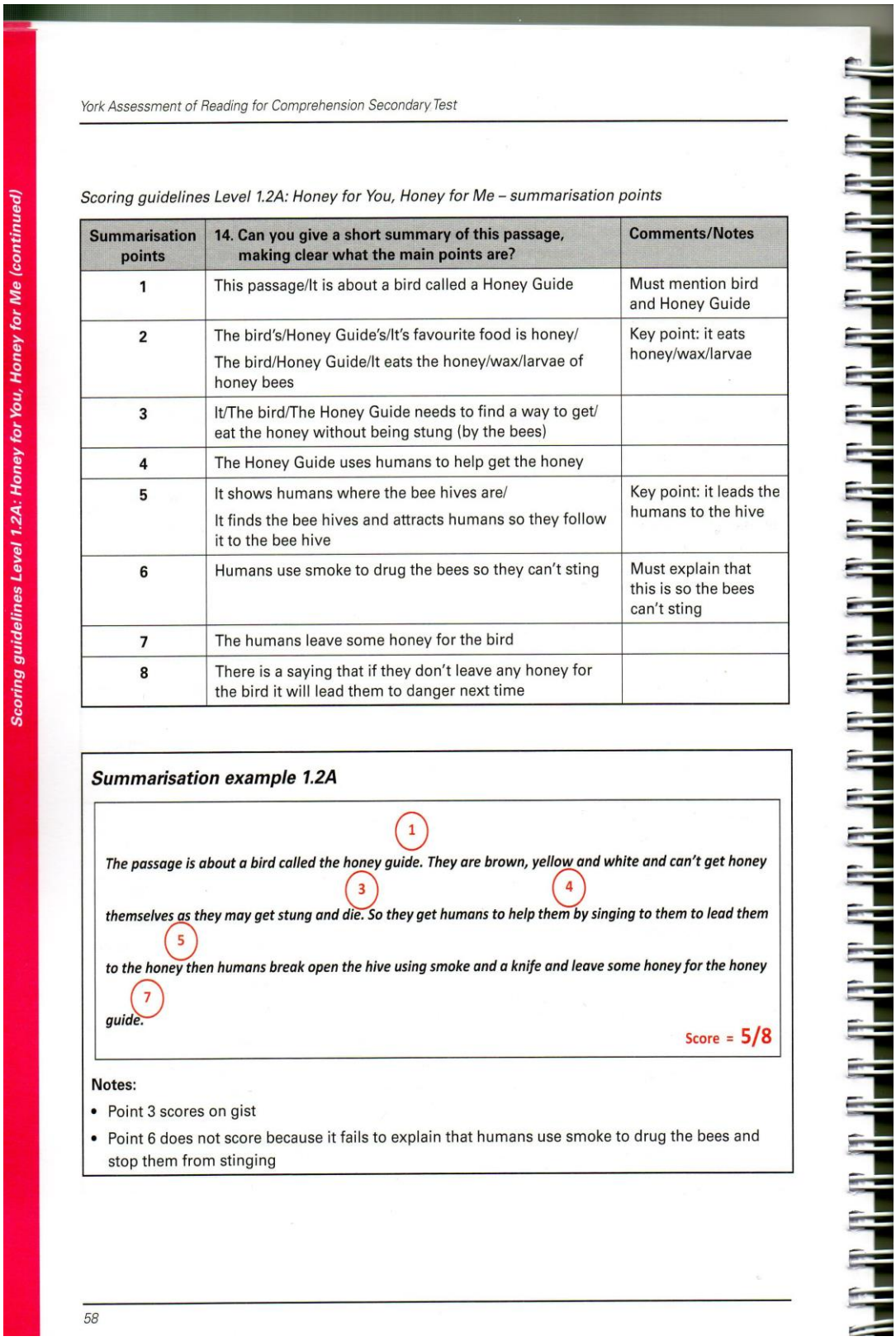
Appendix 4e: Scoring Guidelines to the Reading Comprehension Questions of Passage A

Question	Student's response	Score
1. Suggest three reasons why the bird is called a Honey Guide?	(i) It guides cenedies (people) to honey/ it finds bees' nest for people (ii) It eats honey/ bees wax/ bees eggs/ favourite food is honey/ likes honey (iii) The colour of its chest/ the splash of yellow on its chest (All three answers are required for 1 mark)	1 0 NR
2. What three colours are its feathers?	(i) Brown (ii) White (iii) Yellow/ light yellow (All three answers are required for 1 mark)	1 0 NR
3. Why does the Honey Guide look colourless from a distance?	Because you can only see the brown feathers Because you can only see brown colours on it Because it looks brown From a distance you can only see brown feathers, but up close you can see the yellow	1 0 NR
4. Why are the bees so dangerous to the Honey Guide?	Because of the bees' toroko (sting)	1 0 NR

Appendix 4f: Reading Summarisation Question of Passage A

5. Can you give a short summary of this passage, making clear what the main events are?

Appendix 4g: Scoring Guidelines to the Reading Summarisation Question of Passage A



Scoring guidelines Level 1.2A: Honey for You, Honey for Me – summarisation points

Summarisation points	14. Can you give a short summary of this passage, making clear what the main points are?	Comments/Notes
1	This passage/It is about a bird called a Honey Guide	Must mention bird and Honey Guide
2	The bird's/Honey Guide's/It's favourite food is honey/ The bird/Honey Guide/It eats the honey/wax/larvae of honey bees	Key point: it eats honey/wax/larvae
3	It/The bird/The Honey Guide needs to find a way to get/eat the honey without being stung (by the bees)	
4	The Honey Guide uses humans to help get the honey	
5	It shows humans where the bee hives are/ It finds the bee hives and attracts humans so they follow it to the bee hive	Key point: it leads the humans to the hive
6	Humans use smoke to drug the bees so they can't sting	Must explain that this is so the bees can't sting
7	The humans leave some honey for the bird	
8	There is a saying that if they don't leave any honey for the bird it will lead them to danger next time	

Summarisation example 1.2A

The passage is about a bird called the honey guide (1). They are brown, yellow and white and can't get honey themselves as they may get stung and die (3). So they get humans to help them by singing (4) to them to lead them to the honey (5) then humans break open the hive using smoke and a knife and leave some honey for the honey guide (7).

Score = 5/8

Notes:

- Point 3 scores on gist
- Point 6 does not score because it fails to explain that humans use smoke to drug the bees and stop them from stinging

Appendix 4h: The York Assessment of Reading for Comprehension (Passage B)

River Girl

As a young man, I walked this road many times. It is the road that connects my small **pocoko** to the river. Many, many moons ago, my four brothers and I took turns to make this **sataca**. I was stronger then, and faster; the long, dry road did not seem so hard to me. Setting off at sunrise, I was able to fill my bucket with the cool fresh water by midday. The **sataca** back from the river to my **pocoko** certainly took longer, but I always made it with a smile on my face.

There was a girl you see. She lived in the next **pocoko** to my own. Each day she came from her **pocoko** to the river for water as I did. We met often, accidentally at first, and shared glances across the running water. Soon we began meeting intentionally. I found it hard to think of excuses to take turns for my brothers, but she was worth it. We were in love.

One day strangers entered the **pocoko**; they carried tools and machines for digging and building. As quickly as they came, they went and in their place they left a large well. We did not need to go to the river to get the water anymore; the water came to our **pocoko**. I was happy for the people who lived in my **pocoko**, but a ring of sadness vibrated in my heart.

For a while I continued to make the **sataca** to the river. The ready supply of water meant that more crops could be grown, and more men were needed to tend them. Opportunities became less and less. I suppose my girl lost hope. I went to look for her whenever I could, but our paths did not cross again. I considered visiting her **pocoko** but then I heard that her hand had been promised to somebody else; I would not have been welcome.
Time passed and the sun continued to rise and set with a flexibility I found comforting. I settled down and raised a family of my own. As my own three children grew, so did my crops and life was happy, most of the time.

But soon the light began gradually to disappear. My wife has long since seen her final sunset and now I survey the fields alone. My mind wanders back to the days of my youth; I think about my lost love and wonder whether her life has been as rich as mine. That is why I make this **sataca** now. It will undoubtedly be my last.

Appendix 4i: Reading Comprehension Questions of Passage B

Question	Student's response	Score
1. Which two things does the road connect?		1 0 NR
2. What did the author and his brothers do many years ago? How many brothers did he have? What was he like when he was young?		1 0 NR
3 When did the author set off to collect water from the river? What time did he set off back home?		1 0 NR
4. Why do you think that the River Girl lost hope?		1 0 NR

Appendix 4j: Scoring Guidelines to the Reading Summarisation Question of Passage B

Question	Student's response	Score
1. Which two things does the road connect?	The small pocoko (village) and the river His/ the author's pocoko (village) and the river (Both answers are required for 1 mark)	1 0 NR
2. What did the author and his brothers do many years ago? How many brothers did he have? What was he like when he was young?	They took turns to make a sataca (trip) to the river. Four brothers Strong / fast (All three answers are required for 1 mark)	1 0 NR
3 When did the author set off to collect water from the river? What time did he set off back home?	At sunrise Midday (Both answers are required for 1 mark)	1 0 NR
4. Why do you think that the River Girl lost hope?	She thought he wasn't going to come to the river any more She didn't think that she'd see him again Because he hadn't visited the river for a long time/ he stopped coming to see her	1 0 NR

Appendix 4k: Reading Summarisation Question of Passage B

5. Can you give a short summary of this passage, making clear what the main events are?

Appendix 4l: Scoring Guidelines to the Reading Summarisation Question of Passage B

Scoring guidelines Level 1.1B: River Girl – summarisation points

Summarisation points	14. Can you give a short summary of this passage, making clear what the main events are?	Comments/Notes
1	The author/man when he was young used to go to the river to get water	
2	He used to meet a girl there/at the river *or* He fell in love with a girl he saw at the river	
3	A well was built in the village	
4	He no longer had to fetch water/go to river/ He found it difficult to go to the river (and he didn't see the girl again)	
5	He heard the girl was to marry someone else/ The girl married someone else	
6	He also got married (and had three children)	Must refer to his wife/getting married
7	His wife died long ago	
8	He thinks about the girl	
9	He goes to the river for one last time	Key point: he made the journey for the last time

Summarisation example 1.1B

It's about a man who when he was a boy, fell in love with a girl he saw at the river (2) but then his village got a well (3) so he didn't have to go to the river anymore (4) and he didn't see her again. He got married (6) and grew old but now he is remembering about when he was younger.

Score = 4/9

Notes:

- Point 4 scores on gist
- Point 8 does not score because it does not refer to him remembering about the girl

Appendix 5: A Sample of the interview

Q: How did you find out the meaning of these non-words, 'cenedies', 'toroko', 'pocoko' and 'sataca'?

I recall having seen these words in the texts & having deduced their meanings from the text

Text (1) toroko: associable with the bee & described as dangerous, hence it is the stinger

cenedies: the bird helpers, I predicted they were birds too since they can follow the honey Guide

Text (2) pocoko: village may be - from her pocoko, my pocoko, our pocoko, lived in

Sataca: journey e.g. sataca 'back from the river', make sataca to the river'

Appendix 5 (Cont.)

Q: How did you find out the meaning of these non-words, 'cenedies', 'toroko', 'poçoko' and 'sataca'?

Cenedies → a kind of bird / 1st time * I found the meaning through from context

2nd time * I remembered the meaning only when I saw the text again but not in the exercise

Toroko → can be a part of a bee (body part) / 1st time * I guessed the meaning from context.

or a type of bee

2nd time * I couldn't remember the word until I read the context again.

Appendix 6a: Testing for the possibility of matching non-words to English words

I would be very grateful if you could participate in the task below. It will only take about 5 minutes.

Do you think these words are close to English words?

Please put a tick to indicate how much each word is like an Arabic word on the scale provided.
Please only put ONE tick against each word.

Non-words	Not at all 1	A little 2	Quite 3	A lot 4	Extremely 5
beroro					
bodela					
batoco					
banava					
balado					
bofora					
basati					
beroro					
burota					
cabula					
canilo					
conoro					
comobo					
catila					
cosabi					
cenedi					
conono					
cinito					
ceteri					
conena					
deripa					
dilori					
dogoma					
dotoco					
damana					
dorika					
dufina					
forora					
galena					
gudita					
galapi					
humera					
hapoda					
homolo					
hubara					
homogi					
horoba					
horozo					

Appendix 6a (Cont.)

Non-words	Not at all 1	A little 2	Quite 3	A lot 4	Extremely 5
huraro					
henidi					
jesita					
jarani					
kolena					
latida					
lihora					
lanara					
lapido					
limidi					
mofata					
munido					
menosa					
mosita					
nikalo					
nonana					
nonati					
penita					
palino					
pereto					
pocoko					
parina					
quorata					
raligo					
reseri					
ridoto					
rudaga					
seneli					
sataca					
suderi					
toroko					
tosolo					
venena					
witiro					
wolono					

Thank you very much for your participation☺

Appendix 6b: Testing for the possibility of matching non-words to Arabic words

I would be very grateful if you could participate in the task below. It will only take about 5 minutes.

Do you think these words are close to Arabic words?

Please put a tick to indicate how much each word is like an Arabic word on the scale provided.
Please only put ONE tick against each word.

Non-words	Not at all 1	A little 2	Quite 3	A lot 4	Extremely 5
beroro					
bodela					
batoco					
banava					
balado					
bofora					
basati					
beroro					
burota					
cabula					
canilo					
conoro					
comobo					
catila					
cosabi					
cenedi					
conono					
cinito					
ceteri					
conena					
deripa					
dilori					
dogoma					
dotoco					
damana					
dorika					
dufina					
forora					
galena					
gudita					
galapi					
humera					
hapoda					
homolo					
hubara					
homogi					
horoba					
horozo					

Appendix 6b (Cont.)

Non-words	Not at all 1	A little 2	Quite 3	A lot 4	Extremely 5
huraro					
henidi					
jesita					
jarani					
kolena					
latida					
lihora					
lanara					
lapido					
limidi					
mofata					
munido					
menosa					
mosita					
nikalo					
nonana					
nonati					
penita					
palino					
pereto					
pocoko					
parina					
quorata					
raligo					
reseri					
ridoto					
rudaga					
seneli					
sataca					
suderi					
toroko					
tosolo					
venena					
witiro					
wolono					

Thank you very much for your participation☺

Appendix 7: The item used in the LDT

The item used in the LDT

a) Real word

Passage A	Passage B	Two weeks later A&B	
From the texts			
Splash	Sadness	Splash	Sadness
County	Family	County	Family
Content	Midday	Content	Midday
Colour	Survey	Colour	Survey
Helper	Bucket	Helper	Bucket
problem	Supply	problem	Supply
Not occurring in the texts			
Flower	kitchen	Garden	Consumer
Picture	Forest	Profile	Travel
Bicycle	Poster	Animal	Window
Tissue	Garlic	health	Rainbow
Not occurring in the texts which have a cvcvcv structure			
Employee	Manager	Boundary	Country
Computer	Company	Ability	Agency
potato	Summary	Secondary	Society
primary	tomato	academy	Comedy

b) Non-words

Passage A	Passage B	Two weeks later A&B	
Bataco	Bofora	Cinito	Hafada
Catila	Homolo	Conena	Beroro
Comobo	Kolena	Witiro	Hapoda
Dilori	Dotoco	Tejeja	Dicado
Mofata	Raligo	Rudaga	Hubana
Dorika	Mosita	Homogi	Reseri
Huraro	Limidi	Cosabi	Dufina
Cabula	Gudita	Horoba	Fabaja
Pereto	Nikalo	Wagaki	Conono
Nonana	Munido	Vebege	Burota
Wolono	Suderi	Horozo	Deripa
Ceteri	Tosolo	Conoro	Lagoja
Menosa	Lihora	Seneli	Mecabo
Penita	Latida	Venena	Parina
Damana	Canilo	Ronofo	Nonati
Jesita	Galena	Pefari	Quorata

Appendix 8: Instruction

Instructions

Words will appear individually in the middle of your screen.

You are to decide whether you have seen each word before as quickly and as accurately as possible.

Press 'M' if yes (seen before).
Press 'X' if no (not seen before).

You will see a '*' in the centre of the screen for 20 seconds before the first word is presented. During this time, please place a finger of your left hand on the 'X' key and a finger of your right hand on the 'M' key, so that you are ready to start when the '*' disappears.

You will have 6 practice trials to familiarise you with the procedure

Press the 'Space Bar' when you are ready to start.

Appendix 9: Information sheet

I am a PhD student in DELAL at Reading University. I will be running over the next few days my experiment, and I encourage you to participate for a sum of £10. The study is focus on reading comprehension in Saudi learners of English and I need a control group of L1 English speakers for the material that I will be using. The participants will need to meet with me on two different occasions, as follows:

Day 1: You will be asked to complete the following (35 minutes):

1- A multiple-choice test.
2- A short reading passages with 4 comprehension questions.
3- Another task based on computer.
1- Another short reading passage with 4 comprehension questions.
2- Another task based on computer.

Two weeks later: You will be asked to complete the following (20 minutes):

1- A learning exercise.
2- Another task based on computer.

In one of the tasks, I am using an interesting experimental technique that is employed in psycholinguistics studies, so I hope that you will find the experience as a participant informative as well as enjoyable.

I strongly encourage you to take part and to contribute to this interesting study. If you are keen to do so, please contact me on: r.i.m.alkhudiry@pgr.reading.ac.uk

My mobile number: 07526903440

Thank you very much for your participation☺

Reham Alkhudairy

Appendix 10: Ethics Committee

School of Literature and Languages
Department of English Language and Applied Linguistics

University of Reading

ETHICS COMMITTEE

Project Submission

Note All sections of this form should be completed.

Principal Investigator (Supervisor): Prof. Jeanine Treffers-Daller and Dr. Jacqueline Laws

Student name: Reham Ibrahim Alkhudiry

Department: English Language and Applied Linguistics

Title of Project: Investigating the Relationship between Vocabulary Knowledge and Reading Comprehension Performance: A Study in Saudi Arabia.
Proposed starting date: November 2014

Number of participants that you require consent from (approximate):15-20

I confirm that to the best of my knowledge I have made known all information relevant to the Ethics and Research Committee and I undertake to inform the Committee of any such information which subsequently becomes available whether before or after the research has begun.

I confirm that a list of the names and addresses of the participants in this project will be compiled and that this, together with a copy of the Consent Form, will be retained. All copies of the Consent Forms will be submitted with a copy of the dissertation.

Signed:

…… ……(Supervisor) Date: 22nd October 2014

Reham Alkhudiry……(Student) Date: 22nd October 2014

Appendix 10 (Cont.)

PROJECT DESCRIPTION

this main study.

can be selected for the main experiment.

a Part 3 Option module during the break.

thesis. Data will remain confidential and, apart from myself, only my supervisors will have access to it.

Appendix 10 (Cont.)

University of Reading

Department of English Language and Applied Linguistics

Alkhudiry
@live.reading.ac.uk

HumSS Building
The University of Reading
Whiteknights, PO Box 218
ReadingRG6 6AA

effers-Daller
8-378 2690
-daller@reading.ac.uk

Phone *01183788141*
Email appling@reading.ac.uk

aws
8-378 7460
@reading.ac.uk

Consent Form

Project title: Investigating the Relationship between Vocabulary Knowledge and Reading Comprehension Performance: A Study in Saudi Arabia.

I have read and had explained to me by Reham Ibrahim Alkhudiry the Information Sheet relating to this project.

I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix 10 (Cont)

INFORMATION SHEET

constructed non-words resemble real English words.

at any time, and can inform you about the results of the study once data collection is complete.

This project has been subject to ethical review by the School Ethics and Research Committee, and has been allowed to proceed under the exceptions procedure as outlined in paragraph 6 of the University's *Notes for Guidance* on research ethics.

address above or by email at j.c.treffers-daller@reading.ac.uk or j.v.laws@reading.ac.uk

Signed

Appendix 10 (Cont.)

School of Literature and Languages
Department of English Language and Applied Linguistics

ETHICS COMMITTEE

Project Submission

Note All sections of this form should be completed.

Principal Investigator (Supervisor): Prof. Jeanine Treffers-Daller and Dr. Jacqueline Laws

Student name: Reham Ibrahim Alkhudiry

Department: English Language and Applied Linguistics

Title of Project: Investigating the Relationship between Vocabulary Knowledge and Reading Comprehension Performance: A Study in Saudi Arabia.

Proposed starting date: September 2014

Number of participants that you require consent from (approximate): 60

I confirm that to the best of my knowledge I have made known all information relevant to the Ethics and Research Committee and I undertake to inform the Committee of any such information which subsequently becomes available whether before or after the research has begun.

I confirm that a list of the names and addresses of the participants in this project will be compiled and that this, together with a copy of the Consent Form, will be retained. All copies of the Consent Forms will be submitted with a copy of the dissertation.

Signed:

……(Supervisor) Date: 18th July 2014

Reham Alkhudiry……(Student) Date: July 2014

Appendix 10 (Cont.)

Project Description

The current study aims to investigate the relationship between depth and breadth of vocabulary knowledge and reading comprehension; how reading can contribute to vocabulary retaining, and how depth and breadth can contribute to reading comprehension performance in L1 Arabic learners of English. This study aims to explore those aspects of words which L2 learners can acquire and retain through reading comprehension (meaning, form and use of words). An additional aim is to identify the extent in which the learners' vocabulary size or vocabulary depth can explain variance in reading comprehension.

This investigation will seek to answer the following research questions:

RQ1: which aspects of word knowledge (form, meaning and use) of new words encountered in a reading comprehension task will L2 readers retain?

RQ2: To what extent can vocabulary size and vocabulary depth explain variance in reading comprehension?

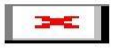
In order to answer RQ1, two measures will be used to assess vocabulary depth more fully by using the Vocabulary Knowledge Scale (Wesche & Paribakht, 1996) developed by (Brown, 2008) and a Lexical Decision Task. The York Assessment Reading Comprehension (YARC) Secondary Test, published by (Stothard, Snowling, Clarke, Barmby & Hulme, 2010), will be used to assess reading comprehension performance. In order to answer RQ2, the Vocabulary Size Test, developed by Nation & Beglar (2007) will be used to measure participants' vocabulary size, and the two depth measures mentioned above will be also used. The York Assessment Reading Comprehension (YARC) Secondary Test will be also introduced to assess reading comprehension performance. It is predicted that the form of the new words will be retained better than their meaning or aspects of use (e.g., collocation) and vocabulary depth will explain more variance in reading comprehension than vocabulary size.

Participants will be volunteers from Qassim University and Reading University. There will be a total of sixty participants and they will be balanced for gender; thirty Saudi Arabic (fifteen females, fifteen males) learners of English from Qassim University and thirty native English speakers (fifteen females, fifteen males) from undergraduate classes at the University of Reading will be asked to participate in the present study. The mean age of all will be twenty years (range: 18-25 years).

The data will be stored on a password protected laptop computer and destroyed immediately after completion of the thesis. Data will remain confidential and, apart from myself, only my supervisors will have access to it.

Appendix 10 (Cont)

School of Literature and Languages
Department of English Language and Applied Linguistics



ETHICS COMMITTEE

Consent Form

Project title: Investigating the Relationship between Vocabulary Knowledge and Reading Comprehension Performance: A Study in Saudi Arabia.

I have read and had explained to me by Reham Ibrahim Alkhudiry the Information Sheet relating to this project.

I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix 10 (Cont.)

Researcher:
Reham Ibrahim Alkhudiry
Email: nc801345@live.reading.ac.uk

Supervisor:
Prof. Jeanine Treffers-Daller
Phone: 00-44-118-378 2690
Email: j.c.treffers-daller@reading.ac.uk
Dr. Jacqueline Laws
Phone: 00-44-118-378 7460
Email: j.v.laws@reading.ac.uk

Department of English Language and Applied Linguistics

HumSS Building
The University of Reading
Whiteknights, PO Box 218
Reading RG6 6AA

Phone 01183788141
Email appling@reading.ac.uk

INFORMATION SHEET

You are being invited to take part in a research study investigating vocabulary knowledge and reading comprehension performance. Firstly, you will be asked to complete a brief questionnaire with information about your educational background, age and interests. Participation in this study involves completing various tasks over three sessions:

Day 1: You will be asked to complete the following:
1) Read a list of 70 words aloud in English (5 minutes)
2) A short test that indicates what vocabulary you are familiar with (25 minutes)

Day 2: You will be asked to complete the following:
1) A short reading passage with 4 comprehension questions (30 minutes)
2) A word recognition task of 20 words (30 minutes)

Day 3: You will be asked to complete the following:
1) Another short reading passage with 4 comprehension questions (30 minutes)
2) Another word recognition task of 20 words (30 minutes)

Two weeks later: You will be asked to complete the following:
1) An exercise that requires you to indicate how well you know a word (30 minutes)
2) Another word recognition task of 20 words (30 minutes)

You may decide to stop being a part of the research study at any time without explanation. You have the right to omit or refuse to answer to any question that is asked of you. You have the right to ask any questions about the procedures answered (unless answering these questions would interfere with the study's outcome). If you have any questions as a result of reading this information sheet, you should ask the researcher before the study begins. The data we collect do not contain any personal information about you. The researcher will be glad to answer your questions about this study at any time, and can inform you about the results of the study once data collection is complete.

This project has been subject to ethical review by the School Ethics and Research Committee, and has been allowed to proceed under the exceptions procedure as outlined in paragraph 6 of the University's *Notes for Guidance* on research ethics.

If you have any queries or wish to clarify anything about the study, please feel free to contact my supervisor at the address above or by email at j.c.treffers-daller@reading.ac.uk or j.v.laws@reading.ac.uk

Signed

Appendix 10 (Cont.)

School of Literature and Languages
Department of English Language and Applied Linguistics

Please go through the checklist below and make sure all the boxes can be ticked before submitting your ethics document. Enclose a copy of the completed checklist to your ethics document.

A. Does your ethics document include

a Project Submission, Project Description, Information Sheet & Consent Form?	☐

B. In your Information Sheet for the participants, have you mentioned the following points? Put a tick in if you have.

The data will be securely kept on a password-protected computer **or** in a locked drawer.	☐
Only the researcher and their supervisors will have access to the data.	☐
The data will be used for academic purposes only.	☐
The data will be anonymous **or** pseudonyms will be used.	☐
The data will be destroyed immediately after the completion of the dissertation.	☐
The participants' privacy and confidentiality will be carefully observed.	☐
The participants have the right to withdraw from the study at any time they wish to.	☐

C. Ensure you have done all the necessary checks.

Have you used the University of Reading logo on all the ethics document pages?	
Have you checked your ethics documents with your supervisor?	
Have you and your supervisor signed the finalised ethics documents?	
Have you included a copy of the first page of your questionnaire, if you are using one?	N/A

D. Only for those who will be working at schools with children and if the school has required for a CRB check.

Have you submitted a copy of your CRB check?	N/A

Appendix 11a: Result of study 1 (Data Distribution)

Data distribution

		TSVST	TSRC	TSSQ	TSVKS	TSLDT	TSTWLDT
N		58	58	58	58	54	54
Normal Parameters[a,b]	Mean	56.5172	5.8621	11.7414	12.4310	59.1667	59.1667
	Std. Deviation	19.32659	1.30382	2.90529	1.51150	3.86933	3.06379
Most Extreme ferences	Absolute	.224	.160	.133	.423	.196	.219
	Positive	.133	.160	.088	.423	.142	.087
	Negative	-.224	-.136	-.133	-.302	-.196	-.219
Kolmogorov-Smirnov Z		1.704	1.215	1.014	3.218	1.443	1.610
Asymp. Sig. (2-tailed)		.006	.104	.256	.000	.031	.011

Appendix 11b: Descriptive statistics of the VKS, LDT (IPT and DPT) target non-word scores for learners of English

Statistics

		TSVKSnw	TSLDTNW	TSTWLDTNW
N	Valid	28	24	24
	Missing	0	4	4

TSVKSnw

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	24	85.7	85.7	85.7
	1.00	1	3.6	3.6	89.3
	2.00	1	3.6	3.6	92.9
	4.00	2	7.1	7.1	100.0
	Total	28	100.0	100.0	

TSLDTNW

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1	3.6	4.2	4.2
	2.00	6	21.4	25.0	29.2
	3.00	9	32.1	37.5	66.7
	4.00	8	28.6	33.3	100.0
	Total	24	85.7	100.0	
Missing	System	4	14.3		
Total		28	100.0		

TSTWLDTNW

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1	3.6	4.2	4.2
	1.00	3	10.7	12.5	16.7
	2.00	6	21.4	25.0	41.7
	3.00	10	35.7	41.7	83.3
	4.00	4	14.3	16.7	100.0
	Total	24	85.7	100.0	
Missing	System	4	14.3		
Total		28	100.0		

Appendix 11c: The association between RCQ scores and (VKS target non-word scores) for L1 Arabic learners of English

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.444[a]	.197	.167	1.00439

a. Predictors: (Constant), TSRC

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.450	1	6.450	6.394	.018[b]
	Residual	26.229	26	1.009		
	Total	32.679	27			

a. Dependent Variable: TSVKSnw

b. Predictors: (Constant), TSRC

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.635	1.213		-2.173	.039
	TSRC	.479	.189	.444	2.529	.018

a. Dependent Variable: TSVKSnw

Appendix 11d: The association between RSQ scores and (VKS target non-word scores) for L1 Arabic learners of English

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.434[a]	.189	.157	1.00981

a. Predictors: (Constant), TSSQ

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.166	1	6.166	6.047	.021[b]
	Residual	26.512	26	1.020		
	Total	32.679	27			

a. Dependent Variable: TSVKSnw

b. Predictors: (Constant), TSSQ

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.605	.835		-1.923	.065
	TSSQ	.173	.070	.434	2.459	.021

a. Dependent Variable: TSVKSnw

Appendix 12: Vocabulary Knowledge Scale (Chapter 4)

Name:.....

A= I know what this word/ phrase means and I can use it in a sentence.
B= I know what this word/ phrase means, but I am not sure how to use it.
C= I've seen this word/ phrase before, but I don't know what it means.
D= I've never seen this word/ phrase before.

	A	B	C	D
beauty				
cenedies				
minute				
bicycle				
toroko				
candle				

Name:.....

A= I know what this word/ phrase means and I can use it in a sentence.
B= I know what this word/ phrase means, but I am not sure how to use it.
C= I've seen this word/ phrase before, but I don't know what it means.

D= I've never seen this word/ phrase before.

	A	B	C	D
door				
pocoko				
weather				
tissue				
sataca				
garden				

Appendix 13a: The York Assessment Reading for Comprehension (Passage A)

Name _____ Date_____

Honey for You, Honey for Me

In Southern Africa there is a bird called the Honey Guide. It is a small bird with a long pink nose. Its favourite food is honey. From a distance, the Honey Guide looks colourless and brown, but up close you can see a splash of light yellow on its white chest feathers. The bird looks a little as if it has just enjoyed a meal of golden honey, and been none too careful about its table manners! However, the Honey Guide gets its name not just from the colour of its chest; it is very well adapted to feeding on the contents of bees' nests. It eats the bees' honey, their eggs and even the bee's wax. In fact, they are one of the only birds that can eat wax. The Honey Guide is what you might call a bee specialist.

It does, however, have one major problem: the bees' **toroko**. The Honey Guide is not a big bird, and a bee **toroko** can be very dangerous, or even kill it so the bird has to find a way to enter the bees' nest and avoid a bad **toroko**.

The Honey Guide has developed a very beautiful solution to the problem. It uses **cenedies**. The Honey Guide searches around its county in the African green lands until it finds a likely-looking bees' nest. When it has found one, it flies off to find some helpers.

The bird attracts the attention of **cenedies** with a non-stop song with short sounds. Once the bird is sure that it has their full attention, it begins to fly off towards the bees' nest, stopping along the way to check that the **cenedies** are following. As the bird flies, it fans its tail out wide, to make it easier to see. When the bird and its helpers reach the bees' nest, the bird calls again to let them know that they have arrived.

Cenedies are well fitted to deal with a nest of angry bees. Using smoke, they drug the bees so they will not be hurt, then break open the nest with a knife to get at the honey pieces. The **cenedies** take as much of the honey as they want, but always leave some for the Honey Guide. In this way the bird has its meal without risk of a harmful **toroko**, and the **cenedies** get help finding hard-to-find bees' nests.

There is a saying among some families that, if the **cenedies** are hungry and do not leave a gift of honey for the Honey Guide, the bird will get angry. If that happens, they say, the next time the bird will guide the **cenedies** not to a bees' nest, but to a lion, an elephant or a harmful snake. It does not pay to be thankless for help, even when your helper is a little bird!

Vocabulary glosses:

cenedies	n. عوائل
	n. مدربسين
	n. أطفال
	n. أشخاص
toroko	n. عش
	n. لسعة
	n. خلية
	N.عسل

Appendix 13b: Multiple-choice Comprehension Questions for Passage A

Reading Comprehension Questions

Please tick the correct option to answer the following questions

Question	Student's response	Score
1. Suggest one reason why the bird is called a Honey Guide?	A. It finds flowers with nectar for people. B. It protects the bees from other animals. C. It helps people to get honey. D. It can distinguish different types of honey from each other.	**1 0 NR**
4. Why are the bees so dangerous to the Honey Guide?	A. Because the bees can sting the bird. B. Because the bees can damage the bird's nest. C. Because the honey bee is not a big bird. D. Because the bees' wax is poisonous for the bird.	**1 0 NR**

Appendix 13c: The York Assessment Reading for Comprehension (Passage B)

Name: **Date:**

River Girl

As a young man, I walked this road many times. It is the road that connects my small **pocoko** to the river. Many, many moons ago, my four brothers and I took turns to make this **sataca**. I was stronger then, and faster; the long, dry road did not seem so hard to me. Setting off at sunrise, I was able to fill my bucket with the cool fresh water by midday. The **sataca** back from the river to my **pocoko** certainly took longer, but I always made it with a smile on my face.

There was a girl you see. She lived in the next **pocoko** to my own. Each day she came from her **pocoko** to the river for water as I did. We met often, accidentally at first, and shared glances across the running water. Soon we began meeting intentionally. I found it hard to think of excuses to take turns for my brothers, but she was worth it. We were in love.

One day strangers entered the **pocoko**; they carried tools and machines for digging and building. As quickly as they came, they went and in their place they left a large well. We did not need to go to the river to get the water anymore; the water came to our **pocoko**. I was happy for the people who lived in my **pocoko**, but a ring of sadness vibrated in my heart.

For a while I continued to make the **sataca** to the river. The ready supply of water meant that more crops could be grown, and more men were needed to tend them. Opportunities became less and less. I suppose my girl lost hope. I went to look for her whenever I could, but our paths did not cross again. I considered visiting her **pocoko** but then I heard that her hand had been promised to somebody else; I would not have been welcome.

Time passed and the sun continued to rise and set with a flexibility I found comforting. I settled down and raised a family of my own. As my own three children grew, so did my crops and life was happy, most of the time.

But soon the light began gradually to disappear. My wife has long since seen her final sunset and now I survey the fields alone. My mind wanders back to the days of my youth; I think about my lost love and wonder whether her life has been as rich as mine. That is why I make this **sataca** now. It will undoubtedly be my last.

Vocabulary glosses:	
sataca	n. القطار على رحلة
	n. رحلة القارب على
	n. الاقدام على رحلة

Vocabulary glosses:	
pocoko	n. ضاحية
	n. مقاطعة
	n. مدينة مركز

Appendix 13d: Multiple-choice Comprehension Questions for Passage B

Reading Comprehension Questions

Please tick the correct option to answer the following questions

Question	Student's response	Score
1. Which two things does the road connect?	A. The castle and the river. B. The suburb and the river. C. The train station and the river. D. The restaurant and the river	**1 0 NR**
2. What did the author and his brothers do many years ago?	A. They went fishing along the river. B. They herded sheep around the river. C. They built a small boat beside the river. D. They took turns to make a trip to the river	**1 0 NR**

Appendix 14a: Elaboration Task with target non-words from Passage A

Elaboration Task

Some of the following sentences contain words you have learned in the reading task. In doing the task below please consider them as normal English words. How appropriate are the following sentences? Give your answer on the following scale (0= completely inappropriate and 6 = completely appropriate). Please explain why?

Examples:

The apple ate a sandwich : 0 (because apples can't eat anything)

Yesterday I goed home at 6 o'clock: 0 (because "goed" should be "went")

John lives in London: 6 (because John is a human being and human beings can live in London)

	Sentences	0	1	2	3	4	5	6	Explain why?
1	The house is big								
2	The cenedies who lived in my city were friendly								
3	I will drank milk with my dinner (grammatically incorrect)								
4	20 cenedies were at the party								
5	She felt nervous before the speech								
6	Some cenedies watched television for a long time								
7	The house seeming clean (grammatically incorrect)								
8	The feathers of cenedies are very thick (semantically anomalous)								
9	Cenedies lived on my farm for two years								
10	The house is fat (semantically anomalous)								
11	I am met many cenedies last night (grammatically incorrect)								
12	A manager should always be honest with his employees								
13	The car walked home (semantically anomalous)								
14	The three legs of cenedies were long (semantically anomalous)								
15	The house is beautiful								
16	The cenedies lovely (grammatically incorrect)								

Appendix 14a (Cont.)

	Sentences	0	1	2	3	4	5	6	Explain why?
1	Mary walked to school in the morning								
2	After getting a toroko from a bee she went to the hospital straightaway								
3	John eats juice for dinner everyday (semantically anomalous)								
4	The toroko is very harmful								
5	The sea water is very cold								
6	She have suffered from the bee's toroko (grammatically incorrect)								
7	The table is unhappy (semantically anomalous)								
8	The toroko was painful for several hours								
9	The table round (grammatically incorrect)								
10	Toroko is made of glass (semantically anomalous)								
11	The table is unstable								
12	The bees' toroko dangerous (grammatically incorrect)								
13	I work late last night (grammatically incorrect)								
14	The toroko was nearly fatal								
15	The table is green								
16	The toroko was angry (semantically anomalous)								

Thank you for your participation

Appendix 14b: **Elaboration Task (distractor)**

Elaboration Task

Some of the following sentences contain words you have learned in the reading task. In doing the task below please consider them as normal English words. How appropriate are the following sentences? Give your answer on the following scale (0= completely inappropriate and 6 = completely appropriate). Please explain why?

Examples:

The apple ate a sandwich : 0 (because apples can't eat anything)

Yesterday I goed home at 6 o'clock: 0 (because "goed" should be "went")

John lives in London: 6 (because John is a human being and human beings can live in London)

	Sentences	0	1	2	3	4	5	6	Explain why?
1	He spent the day drawing in the garden								
2	The **light** dropped from the ceiling								
3	The cat drape herself on the **sofa** (grammatically incorrect)								
4	I managed to carry the **box** upstairs								
5	Mary opened the car **door**								
6	The flower played football for an hour (semantically anomalous)								
7	He have painted a picture of the **garden** (grammatically incorrect)								
8	The light is running (semantically anomalous)								
9	He moved his **hand** slightly to the right								
10	The car walked home (semantically anomalous)								
11	The light broken (grammatically incorrect)								
12	We provided the room with an electrical **heater**								
13	The garden worked very hard last night (semantically anomalous)								
14	The light is shining on the window								
15	The beautiful **garden** attracted many people								
16	The pretty girl home (grammatically incorrect)								

Appendix 14b (Cont.)

	Sentences	0	1	2	3	4	5	6	Explain why?
1	I am reading a good **book** on economics								
2	I bought this **car** very cheaply								
3	The door was happy (semantically anomalous)								
4	He rides his **bicycle** to work every day								
5	The **snake** moved gradually toward its victim								
6	The car hitting a **tree** (grammatically incorrect)								
7	The book liked to drive very carefully (semantically anomalous)								
8	The **trees** surround my property								
9	I met my old friend in the **street everyday** (grammatically incorrect)								
10	My car does not **eat** certain kinds of m**eat** (semantically anomalous)								
11	He picked up the **book** and started to read								
12	The sauce should reduced to one cup (grammatically incorrect)								
13	He slowed down the **car**								
14	I work late last night (grammatically incorrect)								
15	He studied his reflection in the **mirror**								
16	The stone drinks a coffee everyday (semantically anomalous)								

Thank you for your participation

Appendix 14c: Elaboration Task with target non-words from Passage B

Elaboration Task

Some of the following sentences contain words you have learned in the reading task. In doing the task below please consider them as normal English words. How appropriate are the following sentences? Give your answer on the following scale (0= completely inappropriate and 6 = completely appropriate). Please explain why?

Examples:

The apple ate a sandwich : 0 (because apples can't eat anything)

Yesterday I goed home at 6 o'clock: 0 (because "goed" should be "went")

John lives in London: 6 (because John is a human being and human beings can live in London)

	Sentences	0	1	2	3	4	5	6	Explain why?
1	The house is big								
2	The people who lived in my pocoko were friendly								
3	I will drank milk with my dinner (grammatically incorrect)								
4	The pocoko is small								
5	She felt nervous before the speech								
6	The pocoko watched television for a long time (semantically anomalous)								
7	The house seeming clean (grammatically incorrect)								
8	The pocoko is fast (semantically anomalous)								
9	We lived in our little pocoko for five years								
10	The house is fat (semantically anomalous)								
11	I am visited my pocoko last night (grammatically incorrect)								
12	A manager should always be honest with his employees								
13	The car walked home (semantically anomalous)								
14	The pocoko is quiet								
15	The house is beautiful								
16	The pocoko pretty place (grammatically incorrect)								

Appendix 14c (Cont.)

	Sentences	0	1	2	3	4	5	6	Explain why?
1	Mary walked to school in the morning								
2	After the sataca we went to bed straightaway								
3	John eats juice for dinner everyday (semantically anomalous)								
4	The sataca was long								
5	The sea water is very cold								
6	I has enjoyed my sataca to Europe (grammatically incorrect)								
7	The table is unhappy (semantically anomalous)								
8	I ate two bags of crisps on the sataca last night								
9	The table round (grammatically incorrect)								
10	Sataca is made of glass (semantically anomalous)								
11	The table is unstable								
12	We have some time until we started the sataca (grammatically incorrect)								
13	I work late last night (grammatically incorrect)								
14	The sataca was stressful								
15	The table is green								
16	The sataca was angry (semantically anomalous)								

Thank you for your participation

Appendix 14d: **Elaboration Task (distractor)**

Elaboration Task

Some of the following sentences contain words you have learned in the reading task. In doing the task below please consider them as normal English words.
How appropriate are the following sentences? Give your answer on the following scale (0= completely inappropriate and 6 = completely appropriate). Please explain why?

Examples:

The apple ate a sandwich : 0 (because apples can't eat anything)

Yesterday I goed home at 6 o'clock: 0 (because "goed" should be "went")

John lives in London: 6 (because John is a human being and human beings can live in London)

	Sentences	0	1	2	3	4	5	6	Explain why?
1	He spent the day drawing in the garden								
2	The **light** dropped from the ceiling								
3	The cat drape herself on the **sofa** (grammatically incorrect)								
4	I managed to carry the **box** upstairs								
5	Mary opened the car **door**								
6	The flower played football for an hour (semantically anomalous)								
7	He have painted a picture of the **garden** (grammatically incorrect)								
8	The light is running (semantically anomalous)								
9	He moved his **hand** slightly to the right								
10	The car walked home (semantically anomalous)								
11	The light broken (grammatically incorrect)								
12	We provided the room with an electrical **heater**								
13	The garden worked very hard last night (semantically anomalous)								
14	The light is shining on the window								
15	The beautiful **garden** attracted many people								
16	The pretty girl home (grammatically incorrect)								

Appendix 14d (Cont.)

	Sentences	0	1	2	3	4	5	6	Explain why?
1	I am reading a good **book** on economics								
2	I bought this **car** very cheaply								
3	The door was happy (semantically anomalous)								
4	He rides his **bicycle** to work every day								
5	The **snake** moved gradually toward its victim								
6	The car hitting a **tree** (grammatically incorrect)								
7	The book liked to drive very carefully (semantically anomalous)								
8	The **trees** surround my property								
9	I met my old friend in the **street everyday** (grammatically incorrect)								
10	My car does not **eat** certain kinds of m**eat** (semantically anomalous)								
11	He picked up the **book** and started to read								
12	The sauce should reduced to one cup (grammatically incorrect)								
13	He slowed down the **car**								
14	I work late last night (grammatically incorrect)								
15	He studied his reflection in the **mirror**								
16	The stone drinks a coffee everyday (semantically anomalous)								

Thank you for your participation

Appendix 15a: Working Memory Task (Digit Span Test)

Task 1: I am going to say some numbers. Listen carefully, and when I am done, say them after me.

5 - 1 - 7
6 - 2- 3 - 8
4 - 2 - 8 - 3 - 6
7 - 8 - 6 - 4 - 1 - 3
3 - 7 - 5 - 2 - 1 - 6 – 4
8 - 3 - 1 - 9 - 4 - 6 - 2 - 7

Task 2: Now I am going to say some more numbers, but this time when I stop, I want you to say them backwards. For example, if I say 1 -5-9, what would you say?

2 - 6 - 4
7 - 1 - 3 - 6
4 - 2 - 3 - 8 - 7
1 - 6 - 4 - 3 - 9 - 5
7 - 5 - 1 - 2 - 8 - 4 – 9
1 - 3 - 5 - 2 - 9 - 6 - 4 – 8

Task 3: I am going to say some numbers. Listen carefully, and when I am done, say them after me.

٧-٢-١
٣-٦-٤-٩
٥-٢-٦-٨-١
٤-١-٣-٧-٥-٩
٩-٦-٢-١-٤-٨-٣
٢-٥-٣-٧-١-٩-٤-٦

Task 4: Now I am going to say some more numbers, but this time when I stop, I want you to say them backwards. For example, if I say 1 -5-9, what would you say?

١-٣-٩
٨-٦-٢-٧
٦-٤-٢-٥-١
٢-٥-٨-٣-٤-٧
٥-٨-٢-٦-١-٩-٣
٧-٦-٩-٤-١-٣-٨-٢

Appendix 15b: Results of Digit Span Test

Descriptive statistics for DST categories (an outlier identification procedure)

N	FL1	FL2	BL1	BL2
1	8	7	5	5
2	6	6	5	5
3	7	8	5	4
4	6	7	5	5
5	7	6	5	5
6	8	8	5	4
7	8	7	6	5
8	7	6	6	5
9	8	7	6	5
10	6	8	6	5
11	7	7	6	5
12	8	6	5	5
13	7	7	5	5
14	6	6	4	4
15	6	7	4	4
16	7	8	4	4
17	6	7	5	5
18	8	7	6	5
19	7	6	5	5
20	7	7	5	5
21	8	7	4	4
22	7	6	5	5
23	7	7	6	5
24	8	8	5	5
25	7	7	5	5
26	6	6	5	5
27	7	7	6	5
28	7	6	5	4
29	8	7	5	5
30	6	7	5	5
31	7	8	6	5
32	6	6	5	4
33	7	7	6	5
34	7	7	5	5
35	8	8	5	5
36	6	7	5	5
37	7	6	5	4
38	7	8	6	5
39	8	7	5	5
40	7	6	6	5
mean	7.025	6.9	5.2	4.775
std	0.73	0.71	0.61	0.42
std*2	1.47	1.42	1.22	0.85
mean-std*2	5.56	5.48	3.98	3.93
mean+std*2	8.49	8.32	6.42	5.62

Appendix 15b (Cont.)

Mean accuracy scores and standard deviations on the DST categories based on ET-groups

	(ET- target non-words)		(ET- distractors)		Sig/Diff
	Mean	SD	Mean	SD	
Forward in L1	7.00	0.79	7.05	0.69	NS
Forward in L2	6.90	0.72	6.90	0.72	NS
Backward in L1	5.15	0.67	5.25	0.55	NS
Backward in L2	4.75	0.44	4.80	0.41	NS

Mann-Whitney Test for DST categories based on ET-groups

	DST_FL1	DST_FL2	DST_BL1	DST_BL2
Mann-Whitney U	193.000	200.000	186.000	190.000
Wilcoxon W	403.000	410.000	396.000	400.000
Z	-.205	.000	-.435	-.374
Asymp. Sig. (2-tailed)	.838	1.000	.663	.708
Exact Sig. [2*(1-tailed Sig.)]	.862[b]	1.000[b]	.718[b]	.799[b]

a. Grouping Variable: ET-groups

Appendix 16a: Semantic Priming Task

Block (Passage A)

Practice Trials:

Trial 1: (distractor): wajaki (non-word prime) # society (real word) (Yes)
Trial 2: (distractor): book (real word prime) # padiro (non-word) (NO)
Trial 3: (distractor): desktop (semantically related real word prime) # office (real word) (Yes)
Trial 4: (distractor): nikalo (non-word prime) # suderi (non-word) (NO)
Trail 5: (distractor): Mountain (real word prime) # sogano (non-word) (NO)
Trail 6: (distractor): breast (semantically unrelated real word prime) # shower (real word) (Yes)
Trail 7: (distractor): zotabi (non-word prime) # dogana (non-word) (NO)
Trail 8: (distractor): jatela (non-word prime) # castle (real word) (Yes)
Trail 9: (distractor): cousin (semantically related real word prime) # mother (real word) (Yes)
Trail 10: (distractor): infection (real word prime) # zecabo (non-word) (NO)

Real Priming Task:

Trial 1: (distractor): agency (real word prime) # dufina (non-word) (NO)
Trial 2: (distractor): computer (semantically unrelated real word prime) # animal (real word) (Yes)
Trial 3: (distractor): raligo (non-word prime) # munido (non-word) (NO)
Trial 4: (distractor): academy (semantically unrelated real word prime) # sadness (real word) (Yes)
Trial 5: (distractor): conena (non-word prime) # lapido (non-word) (NO)
Trial 6: (distractor): conoro (non-word prime) # dolphin (real word) (Yes)
Trial 7: (distractor): restaurant (semantically related real word prime) # dinner (real word) (Yes)
Trial 8: (distractor): witiro (non-word prime) # cinito (non-word) (NO)
Trial 9: (distractor): lihora (non-word prime) # gobeto (non-word) (NO)
Trial 10: (distractor): tissue (real word prime) # limidi (non-word) (NO)
Trial 11: (distractor): wolono (non-word prime) # history (real word) (Yes)
Trial 12: (distractor): media (real word prime) # seneli (non-word) (NO)
Trial 13: (distractor): flower (semantically related real word prime) # garden (real word) (Yes)
Trial 14: (distractor): fabaja (non-word prime) # mecabo (non-word) (NO)
Trial 15: (distractor): activity (real word prime) # humera (non-word) (NO)
Trial 16: (distractor): bofora (non-word prime) # tejeja (non-word) (NO)
Trial 17: (distractor): shopping (semantically related real word prime) # market (real word) (Yes)
Trial 18: (distractor): bicycle (real word prime) # kolena (non-word) (NO)
Trial 19: (distractor): bodela (non-word prime) # cosabi (non-word) (NO)
Trial 20: (distractor): tosolo (non-word prime) # comment (real word) (Yes)
Trial 21: (distractor): summary (semantically related real word prime) # survey (real word) (Yes)
Trial 22: (distractor): weather (real word prime) # parina (non-word) (NO)

Appendix 16a (Cont.)

Trial 23: (distractor): airport (semantically unrelated real word prime) # bottle (real word) (Yes)
Trial 24: (distractor): ability(real word prime) # lagoja (non-word) (NO)
Trial 25: (distractor): conono (non-word prime) # vebege (non-word) (NO)
Trial 26: (distractor): mobile (real word prime) # homogi (non-word) (NO)
Trial 27: (distractor): factory (semantically unrelated real word prime) # letter (real word) (Yes)
Trial 28: (distractor): problem (real word prime) # jasita (non-word) (NO)
Trial 29: (distractor): comobo (non-word prime) # sajana (non-word) (NO)
Trial 30: (distractor): ridoto (non-word prime) # window (real word) (Yes)

Primes	Targets
Practice Trials	
wajaki	society
book	padiro
desktop	office
nikalo	suderi
Mountain	sogano
breast	shower
zotabi	dogana
jatela	castle
cousin	mother
infection	zecabo
Real Priming Task	
agency	dufina
computer	animal
raligo	munido
academy	sadness
conena	lapido
conoro	dolphin
restaurant	dinner
witiro	cinito

Appendix 16a (Cont.)

lihora	gobeto
tissue	limidi
wolono	history
media	seneli
flower	garden
fabaja	mecabo
activity	humera
bofora	tejeja
shopping	market
bicycle	kolena
bodela	cosabi
tosolo	comment
summary	survey
weather	parina
airport	bottle
ability	lagoja
conono	vebege
mobile	homogi
factory	letter
problem	jasita
comobo	sajana
ridoto	window

Appendix 16a (Cont.)

Block (Passage B)

Practice Trials:

Trial 1: (distractor): balado (non-word prime) # kitchen (real word) (Yes)
Trial 2: (distractor): Jacket (real word prime) # hubara (non-word) (NO)
Trial 3: (distractor): yellow (semantically related real word prime) # silver (real word) (Yes)
Trial 4: (distractor): catila (non-word prime) # dogoma (non-word) (NO)

Trial 5: (distractor): energy (semantically unrelated real word prime) # paper (real word) (Yes)
Trail 6: (distractor): artist (real word prime) # mojari (non-word) (NO)
Trial 7: (distractor): bijara (non-word prime) # system (real word) (Yes)
Trial 8: (distractor): taniza (non-word prime) # vadepo (non-word) (NO)
Trial 9: (distractor): introduction (semantically related real word prime) # abstract (real word) (Yes)
Trial 10: (distractor): parent (real word prime) # socaza (non-word) (NO)

Real Priming Task:

Trial 1: (distractor): heater (real word prime) # dilori (non-word) (NO)
Trial 2: (distractor): school (semantically related real word prime) # primary (real word) (Yes)
Trial 3: (distractor): nonati (non-word prime) # burota (non-word) (NO)
Trial 4: (distractor): camera (semantically unrelated real word prime) # doctor (real word) (Yes)
Trial 5: (distractor): beroro (non-word prime) # deripa (non-word) (NO)
Trial 6: (distractor): potato (semantically related real word prime) # garlic (real word) (Yes)
Trial 7: (distractor): economy (real word prime) # ravama (non-word) (NO)
Trial 8: (distractor): forora (non-word prime) # dicado (non-word)(NO)
Trial 9: (distractor): machine (real word prime) # gudita (non-word) (NO)
Trial 10: (distractor): message (semantically unrelated real word prime) # nature (real word) (Yes)
Trial 11: (distractor): horozo (non-word prime) # batoco (non-word) (NO)
Trial 12: (distractor): manager (semantically related real word prime) # company (real word) (Yes)
Trial 13: (distractor): pefari (non-word prime) # subject (real word) (Yes)
Trial 14: (distractor): ronofo (non-word prime) # venena (non-word)(NO)
Trial 15: (distractor): dapema (non-word prime) # lajeda (non-word)(NO)
Trial 16: (distractor): picture (real word prime) # mosita (non-word) (NO)
Trial 17: (distractor): purple (semantically related real word prime) # colour (real word) (Yes)
Trial 18: (distractor): exercise (real word prime) # dabona (non-word) (NO)
Trial 19: (distractor): nonana (non-word prime) # pereto (non-word) (NO)
Trial 20: (distractor): dotoco (non-word prime) # hapoda (non-word) (NO)
Trial 21: (distractor): calender(real word prime) # nogava (non-word) (NO)
Trial 22: (distractor): profile (semantically unrelated real word prime) # summer (real word) (Yes)
Trial 23: (distractor): perabo (non-word prime) # butter (real word) (Yes)
Trial 24: (distractor): number (real word prime) # rudaga (non-word) (NO)
Trial 25: (distractor): horoba (non-word prime) # cobano (non-word)(NO)
Trial 26: (distractor): dorika (non-word prime) # carrot (real word) (Yes)

Appendix 16a (Cont.)

Trial 27: (distractor): friend (real word prime) # huraro (non-word) (NO)
Trial 28: (distractor): hafada (non-word prime) # family (real word) (Yes)
Trial 29: (distractor): university (real word prime) # tekalo (non-word) (NO)
Trial 30: (distractor): comedy (semantically unrelated real word prime) # health (real word) (Yes)

Primes	Targets
Practice Trials	
balado	kitchen
jacket	hubara
yellow	silver
catila	dogoma
energy	paper
artist	mojari
bijara	system
taniza	vadepo
introduction	abstract
parent	socaza
Real Priming Task	
heater	dilori
school	primary
nonati	burota
camera	doctor
beroro	deripa
potato	garlic
economy	ravama
forora	dicado
machine	gudita
message	nature
horozo	batoco
manager	company

Appendix 16a (Cont.)

pefari	subject
ronofo	venena
dapema	lajeda
picture	mosita
purple	colour
exercise	dabona
nonana	pereto
dotoco	hapoda
calender	nogava
profile	summer
perabo	butter
number	rudaga
horoba	cobano
dorika	carrot
friend	huraro
hafada	family
university	tekalo
comedy	health

Appendix 16b: Instructions

Instructions

Words will appear in separate trials in the middle of your screen.

Each trial will include two sequential stimuli exposed for a fixed time (word then mask) and in the third exposure you should to decide whether you know the meaning of the word as quickly and as accurately as possible.

Press 'M' if yes (known the meaning).
Press 'X' if no (not known the meaning).

You will see a '*' in the centre of the screen for 20 seconds before the first trial is presented. During this time, please place a finger of your left hand on the 'X' key and a finger of your right hand on the 'M' key, so that you are ready to start when the '*' disappears.

You will have 4 practice trials to familiarise you with the procedure

Press the 'Space Bar' when you are ready to start.

Appendix 17: Information sheet

Information Sheet

I am a PhD student in the Department of English Language & Applied Linguistics at the University of Reading. I will be running over the next few days an experiment, and I strongly encourage you to participate. The study focuses on reading comprehension performance in Saudi learners of English. The participants will need to meet with me on three different occasions, as follows:

Day 1: You will be asked to complete the following (40 minutes):

1- A short task to select the appropriate passage.

1- A short reading passage with 4 comprehension questions.

2- A judgment task

3- A computer-based task.

A week later: You will be asked to complete the following (40 minutes):

1- A computer-based task with a learning exercise.

2- Another short reading passage with 4 comprehension questions.

3- Another judgment task

4- Another computer-based task based.

Another week later: You will be asked to complete the following (10 minutes):

1- A computer-based task

2- A learning exercise.

In one of the tasks, I am using an interesting experimental technique that is employed in psycholinguistics studies, so I hope that you will find the experience as a participant informative as well as enjoyable.

I strongly encourage you to take part and to contribute to this interesting study. If you are keen to do so, please contact me on: r.i.m.alkhudiry@pgr.reading.ac.uk

My mobile number: 07526903440

Thank you very much for your participation☺

Reham Alkhudairy

Appendix 18a: Participants' Consent form

Consent Form

I have read and had explained to me by Reham Ibrahim Alkhudairy the Information Sheet relating to this project.

I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described on the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project at any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

1. Name: ABD ULLAH ALAN GARI

2. Native language(s): Arabic

3. Gender: **(a) female** **(b) male** ✓ **(tick one)**

4. Dept/ year of study: 2015

5. Date of birth: 3/ 5/ 1996 20

6. Age at which I first started learning English (for non-native English speakers): 12

7. Number of years I have attended English classes (for non-native English speakers): 8 13 years

8. Number of years / months I have been living in the UK: 1 year

9. My IELTS scores:
Overall: 5.5 **Listening:** 6 **Reading:** 4.5 **Writing:** 5 **Speaking:** 5.5

Date: 9/3/ 2016
Sign: ABD ~~ULLAH~~

Notice: This information will be kept strictly confidential, and is only to help me with the analysis of the results. Once the results have been analysed, all personal information held by me about you will be deleted.

Appendix 18b (Cont.)

Consent Form

I have read and had explained to me by Reham Ibrahim Alkhudairy the Information Sheet relating to this project.

I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described on the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project at any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

1. Name: Mohanned Ali Abualadel

2. Native language(s): Arabic

3. Gender: (a) female (b) male (tick one)

4. Dept/ year of study: Pre sessional/ since october 2015

5. Date of birth: 14/6/1985 30

6. Age at which I first started learning English (for non-native English speakers): 12 years

7. Number of years I have attended English classes (for non-native English speakers): 18 ~~6 years~~

8. Number of years / months I have been living in the UK: 1 year

9. My IELTS scores:
Overall: 6 **Listening:** 6 **Reading:** 6 **Writing:** 5.5 **Speaking:** 6

Date: 09/03/2016

Sign:

Notice: This information will be kept strictly confidential, and is only to help me with the analysis of the results. Once the results have been analysed, all personal information held by me about you will be deleted.

Appendix 19: Ethics Committee

School of Literature and Languages
Department of English Language and Applied Linguistics

ETHICS COMMITTEE

Project Submission

Note All sections of this form must be completed.

Principal Investigator (Supervisor): Prof. Jeanine Treffers-Daller and Dr. Jacqueline Laws

Student name: Reham Ibrahim AlKhudiry

Department: English language and Applied Linguistics

Title of Project: Investigating the Relationship between Vocabulary Knowledge and Reading Comprehension Performance: A Study in Saudi Arabia.

Proposed starting date: December 2015

Number of participants that you require consent from (approximate): 40

I confirm that to the best of my knowledge the Ethics and Research Committee have been made aware of all relevant information. I undertake to inform the Committee of any such information which subsequently becomes available whether before or after the research has begun.

I confirm that a list of the names and contact details of the participants in this project will be compiled and that this, together with signed Consent Forms, will be retained by the researcher under secure storage. All (or in large sample cases a selection) of the signed copies will be submitted with a copy of the dissertation.

Signed:

……(Supervisor) Date: 23/10/2015

Reham Alkhudiry (Student) Date: 21 Oct 2015

Appendix 19 (Cont.)

Project Description

The current study is a part from my previous study which investigates the relationship between vocabulary knowledge and reading comprehension: how reading can contribute to retain vocabulary, and how depth can contribute to reading comprehension performance in L1 Arabic learners of English. This study is an intervention study aimed at finding out to what extent incidental vocabulary learning from reading is most suitable for this particular group of learners to help them learn new words.

This study is based on the Involvement Load Hypothesis which simply claims that the more learners engage with the words they learn, the better they will retain. The York Assessment Reading Comprehension (YARC) Secondary Test (Stothard, Snowling, Clarke, Barmby & hulme, 2010) will be employed to assess reading comprehension performance. A post reading activity, an elaboration task, is created with different involvement loads to investigate how different components of involvement can affect on vocabulary acquisition and retention. Two measures will also be used to assess vocabulary depth (meaning): the Vocabulary Knowledge Scale (VKS) (Wesche & Paribakht, 1996), which was further developed by (Brown, 2008), and a Semantic Priming Task, which will be developed as part of the current project.

Participants will be volunteers from Reading University. Forty participants, Saudi Arabic learners of English, will be asked to participate in the present study. The mean age of all will be twenty years (range: 18-25 years). The data will be stored on a password protected laptop computer and destroyed immediately after completion of the thesis. Data will remain confidential and, apart from myself, only my supervisors will have access to it.

Appendix 19 (Cont.)

School of Literature and Languages
Department of English Language and Applied Linguistics

ETHICS COMMITTEE

Consent Form

Project title: Investigating the Relationship between Vocabulary Knowledge and Reading Comprehension Performance: A Study in Saudi Arabia.

I have read and had explained to me by Reham Ibrahim Alkhudiry the Information Sheet relating to this project.

I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix 19 (Cont.)

Reham Ibrahim Alkhudiry
Email: nc801345@live.reading.ac.uk
Supervisor:
Prof. Jeanine Treffers-Daller
Phone: 00-44-118-378 2690
Email: j.c.treffers-daller@reading.ac.uk
Dr. Jacqueline Laws
Phone: 00-44-118-378 7460
Email: j.v.laws@reading.ac.uk

Department of English Language and Applied Linguistics

HumSS Building
The University of Reading
Whiteknights, PO Box 218
Reading RG6 6AA

Phone 01183788141
Email appling@reading.ac.uk

INFORMATION SHEET

You are being invited to take part in a research study investigating vocabulary knowledge and reading comprehension performance. Firstly, you will be asked to complete a brief questionnaire with information about your educational background, age and interests. Participation in this study involves completing various tasks over four sessions:

Day 1: You will be asked to complete the following:
1) Read a list of 70 words aloud in English (5 minutes)
Day 2: You will be asked to complete the following:
1) A short reading passage with 4 multiple-choice comprehension questions (15 minutes)
2) A post reading activity (30 minutes)
3) A word recognition task of 36 words (10 minutes)
One week later: You will be asked to complete the following:
1) Another word recognition task of 36 words (10 minutes)
2) An exercise that requires you to indicate how well you know a word (10 minutes)
3) Another short reading passage with 4 multiple-choice comprehension questions (15 minutes)
4) A post reading activity (30 minutes)
5) A word recognition task of 36 words (10 minutes)
Another week later: You will be asked to complete the following:
1) Another word recognition task of 36 words (10 minutes)
2) Another exercise that requires you to indicate how well you know a word (10 minutes)
3) A memory test (5 minutes)

You may decide to stop being a part of the research study at any time without explanation. You have the right to omit or refuse to answer to any question that is asked of you. You have the right to ask any questions about the procedures answered (unless answering these questions would interfere with the study's outcome). If you have any questions as a result of reading this information sheet, you should ask the researcher before the study begins. The data we collect do not contain any personal information about you. The researcher will be glad to answer your questions about this study at any time, and can inform you about the results of the study once data collection is complete.

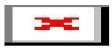
This project has been subject to ethical review by the School Ethics and Research Committee, and has been allowed to proceed under the exceptions procedure as outlined in paragraph 6 of the University's *Notes for Guidance* on research ethics.

If you have any queries or wish to clarify anything about the study, please feel free to contact my supervisor at the address above or by email at j.c.treffers-daller@reading.ac.uk or j.v.laws@reading.ac.uk

Signed

Appendix 19 (Cont.)

School of Literature and Languages
Department of English Language and Applied Linguistics



Please go through the checklist below and make sure all the boxes can be ticked before submitting your ethics document. Enclose a copy of the completed checklist to your ethics document.

A. Does your ethics document include

a Project Submission, Project Description, Information Sheet & Consent Form?	√

B. In your Information Sheet for the participants, have you mentioned the following points? Put a tick in if you have.

The data will be securely kept on a password-protected computer **or** in a locked drawer.	√
Only the researcher and their supervisors will have access to the data.	√
The data will be used for academic purposes only.	√
The data will be anonymous **or** pseudonyms will be used.	√
The data will be destroyed immediately after the completion of the dissertation.	√
The participants' privacy and confidentiality will be carefully observed.	√
The participants have the right to withdraw from the study at any time they wish to.	√

C. Ensure you have done all the necessary checks.

Have you used the University of Reading logo on all the ethics document pages?	√
Have you checked your ethics documents with your supervisor?	√
Have you and your supervisor signed the finalised ethics documents?	√
Have you included a copy of the first page of your questionnaire, if you are using one?	N/A

D. Only for those who will be working at schools with children and if the school has required for a CRB check.

Have you submitted a copy of your CRB check?	N/A

Appendix 20a: Data Distribution

One-Sample Kolmogorov-Smirnov Test

		TSVKS	TSPTA	TSPTB	TSPTOWA	TSPTOWB	TSPTALL	TSDST
N		40	40	40	40	40	40	40
Normal Parameters[a,b]	Mean	21.3750	118.3250	115.0500	113.8500	109.0500	456.2750	24.8000
	Std. Deviation	2.37171	8.35859	8.98988	14.86098	17.86552	32.20924	2.04061
Most Extreme Differences	Absolute	.204	.159	.159	.259	.205	.134	.153
	Positive	.134	.124	.092	.171	.144	.110	.153
	Negative	-.204	-.159	-.159	-.259	-.205	-.134	-.148
Kolmogorov-Smirnov Z		1.290	1.009	1.006	1.638	1.299	.847	.969
Asymp. Sig. (2-tailed)		.072	.261	.264	.009	.068	.470	.304

a. Test distribution is Normal.

b. Calculated from data.

One-Sample Kolmogorov-Smirnov Test

		SWRT	Reading_comp_A	Reading_comp_B	TSELBA	TSElbB	TSElb2
N		40	40	40	20	20	40
Normal Parameters[a,b]	Mean	47.98	1.23	1.800	163.3500	165.6000	149.9000
	Std. Deviation	2.496	.768	.4051	20.17887	14.41271	18.33702
Most Extreme Differences	Absolute	.184	.269	.489	.148	.110	.116
	Positive	.113	.190	.311	.112	.110	.072
	Negative	-.184	-.269	-.489	-.148	-.101	-.116
Kolmogorov-Smirnov Z		1.166	1.699	3.094	.661	.494	.733
Asymp. Sig. (2-tailed)		.132	.006	.000	.776	.968	.656

a. Test distribution is Normal.

b. Calculated from data.

Appendix 20b: Mann-Whitney U-test between ET-A1 and ET-B1

Mann-Whitney U-test between A1 and B1 scores

Ranks

	Learning	N	Mean Rank	Sum of Ranks
TSELTNW	1.00	20	20.33	406.50
	2.00	20	20.68	413.50
	Total	40		

Test Statistics[a]

	TSELTNW
Mann-Whitney U	196.500
Wilcoxon W	406.500
Z	-.095
Asymp. Sig. (2-tailed)	.925
Exact Sig. [2*(1-tailed Sig.)]	.925[b]

a. Grouping Variable: Learning
b. Not corrected for ties.

Mann-Whitney U-test between A2 and B2 scores

Ranks

	Learning	N	Mean Rank	Sum of Ranks
TSElb2	1.00	20	19.85	397.00
	2.00	20	21.15	423.00
	Total	40		

Test Statistics[a]

	TSElb2
Mann-Whitney U	187.000
Wilcoxon W	397.000
Z	-.352
Asymp. Sig. (2-tailed)	.725
Exact Sig. [2*(1-tailed Sig.)]	.738[b]

a. Grouping Variable: Learning
b. Not corrected for ties.

Appendix 20c: The association between ET (A1 and B1), and target non-word scores of VKS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.311[a]	.096	.073	2.16964

a. Predictors: (Constant), TSETNWSENTENCES

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	19.096	1	19.096	4.057	.051[b]
	Residual	178.879	38	4.707		
	Total	197.975	39			

a. Dependent Variable: TSVKSNW

b. Predictors: (Constant), TSETNWSENTENCES

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.244	2.153		.578	.567
	TSETNWSENTENCES	.055	.027	.311	2.014	.051

a. Dependent Variable: TSVKSNW

Appendix 20d: The association between ET (A1 and B1), and target non-word scores of SPT (IPT) for both passages

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.544[a]	.296	.278	3.49677

a. Predictors: (Constant), TSETNWSENTENCES

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	195.759	1	195.759	16.010	.000[b]
	Residual	464.641	38	12.227		
	Total	660.400	39			

a. Dependent Variable: TSSPTNWAB

b. Predictors: (Constant), TSETNWSENTENCES

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.508	3.470		-.723	.474
	TSETNWSENTENCES	.177	.044	.544	4.001	.000

a. Dependent Variable: TSSPTNWAB

Appendix 20e: The association between ET (A1 and B1), and target non-word scores of SPT (DPT) for both passages

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.602[a]	.363	.346	3.05838

a. Predictors: (Constant), TSETNWSENTENCES

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	202.459	1	202.459	21.645	.000[b]
	Residual	355.441	38	9.354		
	Total	557.900	39			

a. Dependent Variable: TSSPTOWNWAB

b. Predictors: (Constant), TSETNWSENTENCES

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.990	3.035		-.326	.746
	TSETNWSENTENCES	.180	.039	.602	4.652	.000

a. Dependent Variable: TSSPTOWNWAB

Appendix 20f: Results of RT on the SPT (IPT and DPT)

Mean RTs and standard deviations on the SPT (IPT and DPT)

	(N=40)				
	SPT/ ET (A1 or B1)		SPT/ ET (A2or B2)		Significant difference
	Mean	SD	Mean	SD	
LDT (IPT)	840.05	194.60	859.70	142.42	NS
LDT (DPT)	807.03	157.75	797.10	123.47	NS

Wilcoxon Signed Ranks Test for SPT (IPT and DPT) between ET- A1 or B1 and ET-A2 or B2

	TSRTETDIS - TSRTETNW	TSOWRTETDIS - TSOWRTETNW
Z	-.632[b]	-.161[c]
Asymp. Sig. (2-tailed)	.528	.872

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.
c. Based on positive ranks.

Means and standard deviations of the RT on the SPT (IPT) categories

(N=40)	SPT/ ET A1or B1		SPT/ ET A2 or B2		Sig/Diff
	Mean	SD	Mean	SD	
Target word primes					
Arabic & English prime (1)	937.30	262.53	884.56	300.44	NS
Arabic & English prime (2)	879.70	583.84	917.13	314.86	NS
Unrelated word prime (3)	890.96	286.92	948.36	290.26	NS
Unrelated non-word prime (4) (Max= 2)	817.29	226.27	883.90	226.56	NS
Distractor item primes					
non-words primed by non-words (5)	860.88	191.05	887.43	196.92	NS
real words primed by semantically related real words (6)	756.51	161.44	789.07	163.99	NS
real words primed by semantically unrelated real words (7)	748.49	145.00	769.81	142.62	NS
real words primed by non-words (8)	780.40	148.19	800.42	149.32	NS
Non-words primed by real words (9)	846.69	187.64	871.96	179.01	NS

Appendix 20f (Cont.)

Wilcoxon Signed Ranks Test for categories of SPT (IPT) between ET- A1 or B1 and ET-A2 or B2

	TSRTETDISTWA - TSRTETNWTWA	TSRTETDISTWE - TSRTETNWTWE	TSRTETDISTWRW - TSRTETNWTWRW	TSRTETDISTWNW - TSRTETNWTWNW
Z	-.228[b]	-1.152[c]	-.400[c]	-1.094[c]
Asymp. Sig. (2-tailed)	.819	.249	.689	.274

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

Means and standard deviations of the RT on the SPT (DPT) categories

(N=40)	SPT/ ET A1or B1		SPT/ ET A2 or B2		Sig/Diff
	Mean	SD	Mean	SD	
Target word primes					
Arabic & English prime (1)	916.55	255.70	808.26	200.14	NS
Arabic & English prime (2)	794.46	237.23	796.79	163.39	NS
Unrelated word prime (3)	855.32	214.32	823.36	261.22	NS
Unrelated non-word prime (4) (Max= 2)	809.87	208.18	813.63	233.78	NS
Distractor item primes					
non-words primed by non-words (5)	822.34	172.10	817.45	140.33	NS
real words primed by semantically related real words(6)	749.13	151.23	766.53	113.92	NS
real words primed by semantically unrelated real words (7)	747.00	161.64	738.34	104.84	NS
Real words primed by non-words (8)	746.46	134.50	775.02	110.45	NS
Non-words primed by real words(9)	809.76	154.70	809.93	132.09	NS

Appendix 20f (Cont.)

Wilcoxon Signed Ranks Test for categories of SPT (DPT) between ET- A1 or B1 and ET-A2 or B2

	TSRTETOWDISTWA - TSRTETOWNWTWA	TSRTETOWDISTWE - TSRTETOWNWTWE	TSRTETOWDISTWRW - TSRTETOWNWTWRW	TSRTETOWDISTWNW - TSRTETOWNWTWNW
Z	-1.564[b]	-.196[b]	-.941[b]	-.487[c]
Asymp. Sig. (2-tailed)	.118	.844	.347	.626

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.