University of Reading

Department of English Language & Applied Linguistics
School of Literature and Languages

Intentional and Incidental Learning of Vocabulary Among L1 Arabic Learners of English Using Word Cards and Graded Readers

Thesis submitted for the degree of Doctor of Philosophy

in Applied Linguistics

Sally Abdullah Alghamdi

May 2018

Declaration

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

S A Alghamdi
May 2018

Abstract

Intentional learning can result in higher retention rates than incidental learning (Elgort, 2011; Horst, Cobb, & Nicolae, 2005; Hulstijn, 2001; Laufer, 2005; Nation, 2013; Prince, 1996). The current study explored the proportion of target words learned by L1 Arabic learners of English intentionally by using word cards and compared it with the proportion of the same target words learned incidentally by participants using graded readers and a control group who received no vocabulary guidance. Word cards techniques have the intentional learning advantages and are more effective than other vocabulary-learning techniques because of their distinctive characteristics such as learning the form and meaning of many vocabulary items in a short time (Hung, 2015; Mondria, 2007; Mondria & Mondria-De Vries, 1994; Nakata, 2008; Nation, 2001; Waring, 2004). Conversely, graded readers, despite the limited number of vocabulary items that can be retained through their use, provide an effective way to learn vocabulary (Day, Omura, & Hiramatsu, 1991; Dupuy & Krashen, 1993; Horst, Cobb, & Meara, 1998; Mason & Krashen, 1997; Pitts, White, & Krashen, 1989). Very few studies have investigated the characteristics of vocabulary retention one to two weeks after the experimental phase, in order to measure longer-term retention or 'long-term' recall of vocabulary (Godwin-Jones, 2010; Pellicer-Sánchez & Schmitt 2010). The current study investigated the effect of spaced repetition techniques, on the retention of the forms and meanings of forty-five low frequency target words (i.e. 6000 onwards according to British National Corpus (BNC)) immediately after each learning session and after a two-week delay. To evaluate learners' vocabulary knowledge, it is important to use different vocabulary measures to investigate the different dimensions and degrees of knowing a word (Daller, Milton, & Treffers-Daller, 2007; Hulstijn, 2010; Milton, 2009; Nation, 2013; Schmitt, 2010). Three tests administrated during the learning phase and a delayed post-test were used to measure vocabulary receptive knowledge by using a Lexical Decision Task (LDT), and productive vocabulary knowledge using a gap-fill test. The overall performance of the word card group was superior to the graded readers and control group in the three testing phases and in the post-test.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisors Prof. Jeanine Treffers-Daller and Dr. Jacqueline Laws for the continuous support of my PhD study, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better advisors and mentors for my PhD study. I would also like to present my gratitude to Dr. Jon Clenton for his help in the beginning of this research by supporting me with feedback and very helpful suggestions.

My sincere thanks also go to the English Literature Program in the Department of Foreign Languages at Albaha University, who provided me an opportunity to join their team and who gave access to the students in this program and to the research facilities. Without their precious support, it would not be possible to conduct this research.

I would like to express my deepest appreciation to all of my family members who have greatly supported me during the process and made it possible; with no doubt that the first are my father Prof. Abdullah Alghamdi and my mother Fawziah Alghamdi for their faith on me and the unlimited love and support they provide all the time. My sincere gratitude also goes to my wonderful brother Saleh and my lovely sisters Sarah, Sana'a and Samia for their unlimited love and spiritual support throughout writing this thesis and in my life in general.

I would like to express my genuine appreciation to my best friend, my soulmate and my rock Dr. Sami Alghamdi as he always provided me with guidance and emotional support from the first day of this journey and in my life in general. I would like to thank my lovely kids, Sarah, Fatimah and Yousuf for their unconditional love and their patience for staying at home so many weekends during the time of conducting this research.

Abstract

Acknowledgments

Table of Contents

List of Appendices

List of Tables

List of Figures

List of Abbreviations

L1	First language
L2	Second language
ESL	English as a second language
WM	Working memory
STM	Short-term memory
BNC	British National Corpus
PLT	Productive Vocabulary Levels Test
RM	Ready Made
FD	Free Design
RMCG	Ready-made word card group
GRG	Graded readers group
CG	Control group
VST	Vocabulary Size Test
VLT	Vocabulary Level Test
PPVT	Peabody Picture Vocabulary Test
TOEFL	Test of English as a Foreign Language
LFP	Lexical Frequency Profile
LDT	Lexical Decision Task
AS	Accuracy Score(s)
RT	Reaction Time
GFT	Gap-fill Test
FWMs	Forward word card movements
BWMs	Backward word card movements
CQs	Comprehension questions

1. Introduction

The research presented in this thesis is an exploration of the effect of intentional vocabulary learning using word cards and the *Leitner system*, and the effect of incidental vocabulary learning using graded readers on Arab ESL learners (see section 2.1. for the definition of *learning* in general and for definition of *intentional vocabulary learning* and *incidental vocabulary learning* specifically). This introductory chapter establishes the importance of vocabulary research, teaching, and learning in general and the importance of intentional and incidental vocabulary learning using learning techniques specifically. It also sheds light on the pedagogic and personal factors that motivated this study. Additionally, the chapter provides the context of the study, a summary of the research undertaken, and finally, it provides an outline of the thesis structure.

1.1 The Importance of Vocabulary

Over the last 20 years, vocabulary research has become a lively subject about which an unprecedented amount research continues to grow. Vocabulary is no longer the "Cinderella subject", but it is a subject that is interrelated to different language skills (Daller et al., 2007). Studying vocabulary encourages the creation of new research designs, new ways of testing, and new methodologies (Daller et al., 2007). It is only recently that vocabulary research has contributed to the mainstream of foreign language teaching (Milton, 2009).

Unfortunately, vocabulary was a neglected subject in applied linguistics research for a long time and this negligence caused a lack in the presence of vocabulary in research and in professional practice. According to Milton (2009), there are three reasons for the underestimation of the importance of vocabulary research. Firstly, the common understanding of vocabulary, outside the field of research, is that vocabulary is merely words that will be learned eventually because of the structural approach of teaching a language. Vocabulary is easy and learners do not need help with its learning. Secondly, it has been the shared belief between teachers, educators, and learners that learning limited vocabulary can be enough for the knowledge of a foreign language. This contrasts with the beliefs of Laufer, (1989), Nation (2013) and Schmitt (2000) that second language learners need to keep accumulating vocabulary while learning a second language to be able to understand written and spoken language. Thirdly,

there is an over-reliance of teachers on the incidental learning of vocabulary by reading or listening because the explicit teaching of vocabulary is time consuming and will not result in high levels of vocabulary learning. A response to these false beliefs about vocabulary follows.

The term vocabulary is an umbrella term for a wide range of concepts. Vocabulary can be used to refer to single words (e.g., *child*) or multiple words or expressions (e.g., *baby, kids, toddler, infant*) that have a fixed meaning. An important concept in the field is that of the lexeme which represents the base meaning of a word (Nunan, 2012, p. 43). Thus, the words *going* and *went* are different forms of one lexeme, namely *go.* A lemma includes the baseword (lexeme) and its most frequent inflections (Milton, 2009, p. 10). Importantly, the part of speech remains the same for all words that comprise the lemma. Thus, *going, went*, and *goes* are different forms of the verb *go*, and *happy, happier*, and *happiest* are different forms of the adjective *happy*. When derivational affixes, such as –able, or un-, are attached to a baseword (as in *workable*, *unworkable*) more complex words are created and these all form different lemmas. These affixes play an important role in modifying meaning (Laws & Ryder, 2014) and they may also change the part of speech of a word. A word family comprises all inflected and derived forms of a baseword. Thus, *workable, worker, working*, and *worked* all belong to the word family *work*. The members of the word family can thus belong to different parts of speech. The distinctions between base words, lemma's, and word families is important if vocabulary is counted in research, for example if one wants to measure vocabulary size. If the units counted are basewords, the result will, therefore, be different from calculations in which lemma's or word families are counted.

Vocabulary is an important component of language. For learners to develop good language comprehension, accumulating a good proportion of vocabulary items is crucial. The shared belief between teachers, educators, and learners that learning limited vocabulary can be enough for the knowledge of a foreign language is not necessarily true. According to Laufer & Ravenhorst-Kalovski (2010) and Nation (2013), readers need to know 95% of the words in a text to obtain a basic understanding. It would be difficult for a learner with limited vocabulary knowledge to read a natural text read by native speakers. Therefore, Nation (2013) suggests that it should be a priority for second language (L2) learners at the early intermediate stages to learn high frequency words.

The average native English language speaker acquires five to ten words per day from age one year to age six through everyday communication (Gleitman & Gleitman, 1994). Children then build on their vocabulary knowledge by learning infrequent words (Gleitman & Gleitman, 1994). Frequency is also related to age of acquisition, in that children learn high frequency words first and very few words from the low frequency word list (Harley, 2008). As native speakers typically add 1000 words to their vocabulary each year during their formative

years it is possible that the number of words known by an educated native speaker is around 20,000 word families (Nation, 2001). It is important to learn high frequency vocabulary first in the early stages of learning the L2; learning low frequency words helps with gaining high frequency vocabulary knowledge (Gleitman & Gleitman, 1994). Teaching and learning high frequency vocabulary and low frequency vocabulary has been a critical subject of many vocabulary research.

The over-reliance of teachers on learning vocabulary incidentally from reading or listening, and the underestimation of the value of explicit teaching and intentional learning to vocabulary is a problem. Fortunately, this problem prompts a wide range of research on incidental vocabulary learning (i.e., learning words as a by-product of reading or listening) and few on intentional vocabulary learning (i.e., learning words through tasks with the specific aim of learning words). Extensive reading, or reading long texts for enjoyment without concentrating too much on meaning, helps learners learn words incidentally, but this depends on them having the opportunity to use the new words, or to remember them (Nation, 2013). However, there is growing evidence that in certain cases and stages of learning, intentional vocabulary learning is more effective than incidental vocabulary learning (Nation, 2013). Incidental vocabulary learning is not quite sufficient for fast and specific vocabulary list learning because learners acquire or retain the meaning of only a small proportion of the new words they encounter in a text. However, the rate of the ability to remember words through intentional learning techniques is more than from incidental learning (Hulstijn, 2003). Therefore, teachers need the knowledge of how to apply intentional and incidental vocabulary learning techniques in the classroom. This knowledge cannot be obtained without research and professional practice.

1.2 Factors Motivating the Study

Pedagogic factors

Nation (2013) suggests that a vocabulary course should involve the explicit teaching of vocabulary and should train learners to learn and study vocabulary intentionally. The powerful impact of explicitly drawing attention to certain features of vocabulary is beneficial. Furthermore, Nation (2013) claims that for effective vocabulary learning and for gaining a range of aspects of knowledge of a vocabulary item, it is important to encounter a vocabulary item in different contexts and through different modalities. Therefore, intentional vocabulary

learning can be combined with incidental vocabulary learning from reading and listening (Cohen, 1987; Nation, 2013).

In applied linguistics, there is an interchangeable relationship between professional practice and good research. It is important to both benefit from vocabulary research by applying it in the classroom and to conduct vocabulary research in the classroom (Nation, 2011). Teachers need support in teaching vocabulary from research. Research should provide an analysis of the design of vocabulary learning techniques to help teachers to design learning vocabulary tasks (Nation, 2013). Furthermore, it would be better for research to challenge the assumption some teachers hold against vocabulary teaching and learning, which they do by over evaluating the importance of grammar (Folse, 2004). Also, research should challenge the assumption some teachers hold against explicit vocabulary teaching and intentional vocabulary learning, which they do by over evaluating the importance of implicit vocabulary teaching and incidental vocabulary learning (Nation, 2013).

Many teachers realize the importance of learning vocabulary and they try to use the methods supported by research. However, in many cases teachers find it difficult to use new vocabulary teaching methods because they need research to simplify the job of teaching and learning vocabulary by providing real and practical examples (Nation, 2013). Teachers should be trained to know how to measure a learner's vocabulary size and what specific principles they should apply when teaching high and low frequency words. This cannot be achieved without the help of research (Nation, 2013).

One technique to learn vocabulary intentionally is spaced repetition using word cards. The vocabulary card technique adheres to student-centred principles, not only because it can be used independently, but also because the cards can be personalized according to the individual's needs and desires (Elgort, 2011). To maximize the benefits of word cards, their use can be repeated and spaced over time by using the *Leitner system* (Mondria & Mondria-De Vries, 1994). The *Leitner system* applies the spaced repetition technique either manually, by using word cards and a box, or automatically, by using a computer (Mondria & Mondria-De Vries, 1994). The card movement is based on algorithms which are becoming more popular in computer programming research (e.g., Amiri, Miller, & Savova, 2017; Reddy, Labutov, Banerjee, & Joachims, 2016).

Furthermore, vocabulary research is important not only to provide information of the suitable techniques to teach and learn vocabulary but also to focus more on how vocabulary knowledge grows over time (Nation, 2013). Also, although there is a wide range of tests to determine vocabulary size, there is a need to focus on how to improve vocabulary size tests and their precise purpose when used by researchers or by teachers (Kremmel & Schmitt, 2016).

Personal experience

From my experience as an English teacher in all female schools and as an English lecturer at Albaha University in Saudi Arabia, I strongly believe that the major weakness in my students' language is the lack of low frequency vocabulary. Now, from my position as a lecturer in the Department of Foreign Languages, I experience the struggle my students have while learning English because of their lack of vocabulary. The Department of Foreign Languages at Albaha University recognizes the importance of devoting significant time to vocabulary learning and teaching. As a result, they require students to attend a vocabulary-building course during the first year of the English Literature program in the Department of Foreign Languages (see Appendix 1 (B)). However, the problem is students are not getting the most out of these classes because the lecturers are using old fashioned ways of teaching vocabulary (i.e., the behaviourist approach to learning focuses on teaching by constant repetition of a task and it depend on receiving feedback from the instructor), or they are depending on the learners self-learning through textbook exercises and lessons. When I applied a word cards technique (without spaced repetition) the students were reluctant to take part without receiving extra points in their assessment. However, at the end of the course most of the students liked this technique and wanted to use it for their free vocabulary learning and for other subjects as well. Although the vocabulary-building course is now a requirement, it is not enough and more attention to vocabulary learning and teaching should be considered.

In another experience of teaching vocabulary in an independent, non-profit organization (International House (I- House), Philadelphia USA) I was given the choice to implement a student-centred learning approach. I applied a vocabulary card technique (without spaced repetition). I created vocabulary card activities which students found complex because they had to find new words, write down the information they discovered, write a story using these new words, and then present them in class for ten minutes. This task made my students aware of the benefit of word cards. They became selective in their word choices, wrote definitions, and tried to create their own examples. They also said that they would use this method for upcoming self-study. From these previous teaching experiences, and from my own experience as a learner who used word cards while studying English at the University of Pennsylvania, I am motivated to find out more about word card techniques as I think they are the best solution for speedy learning of a long list of vocabulary items in a very short period of time.

1.3 Context

This study was conducted at Albaha city in the southwest region of Saudi Arabia. The Department of Foreign Languages and Faculty of Arts & English Language Division at Albaha University provided support and context for the study. I have been teaching in this Department for about two years, therefore the context is familiar to me. Standard Arabic is the language of Saudi people, the main foreign language taught in schools, and spoken outside of schools, is English.

For two years, I have been teaching vocabulary, reading, and grammar subjects to students at the same level as the participants in the study. It is important to note that Arabic is a Semitic language with a unique word structure that is very different from English (Milton, 2016). Understanding the special nature of Arabic, the difficulties facing Arab ESL learners, and considering the participants language proficiency are all considered in this study.

1.4 The Focus of the Research

In the current study a word card technique is applied to examine intentional vocabulary learning and graded readers are applied to examine incidental vocabulary learning. Both techniques are designed to ensure the rehearsal and repetition of a number of target words over a three-week period of learning. The current study explores the potential effects of word cards and the *Leitner system* versus graded readers and planned vocabulary repetition on the retention of Arab ESL learners' vocabulary knowledge (see section 2.4.2 for the definition of *retention*). The study explores the approximate number of repetitions necessary for intentional vocabulary retention and compares this with the approximate number of repetitions necessary for incidental vocabulary retention. Furthermore, it investigates the amount of spacing between repetitions needed for intentional vocabulary learning, and for incidental vocabulary learning in a quantitative analysis. The recognition (i.e., defined in section 2.2.2) and production (i.e., the producing of word forms in speaking or writing to communicate meaning) of target words during three weeks of learning and target word retention growth are analysed quantitatively. The target word characteristics that can cause difficulty for English learners in general, and the target word characteristics that can cause difficulty for Arab ESL learners specifically, are analysed qualitatively. The performances of ready-made word card groups (RMCG), using ready-made word cards, and a graded readers' group (GRG), using the graded readers, are

compared with the performance of a control group (CG) who received no vocabulary teaching or activities.

The current study has important practical implications for vocabulary teaching. Word cards and spaced repetitions are beneficial to learning and can easily fit into an existing educational system because it is viable and does not consume too much time or money (Kang, 2016). Furthermore, teachers want to encourage students in their classes to use techniques that are supported by research (Mondria, 2007; Nation, 2013). However, in many cases teachers find it difficult, or do not wish to use new vocabulary teaching methods, because of the lack of research on vocabulary-techniques (Nation, 2013). More implications of vocabulary teaching and learning is discussed in the concluding chapter of this thesis.

1.5 Thesis Structure

Following this introductory chapter, the Literature Review Chapter shows how the literature has helped refine the rationale for the investigation on which the thesis reports. Thus, it provides a literature review concerning vocabulary knowledge and previous research on intentional and incidental vocabulary learning, the *forgetting curve* and rehearsal techniques, and techniques for supporting vocabulary learning. By the end of this chapter, justification for the study is addressed.

Research methodology is the focus of the Methodology Chapter and it includes the pilot study design, methodology, research questions and hypotheses and the results of pre-tests of the pilot study followed by an explanation of the design of the main study and results of pre-tests. Furthermore, the schedule of the tasks and main study are explained.

The Pilot Study Chapter is the fourth chapter and it presents the pilot study and consists of an evaluation of the research questions. The fifth chapter is the Results Chapter and it presents the results of the study. In this chapter, an analysis of the tests used and responses to each of the research questions are addressed. The majority of Chapter 5 comprises a quantitative analysis of the results, while a minor section is dedicated to the qualitative analysis of the results.

The sixth chapter is the Discussion Chapter and it provides supporting evidence and description of the results. This chapter provides a summary of the results, followed by an indication of the importance of the findings for each of the research questions and what generalizations can be made.

Finally, the seventh chapter is the Conclusion Chapter and it summarises the findings and considers pedagogical implications by drawing on the findings reported in the Results Chapter

and on the discussion seen in the Discussion Chapter. Concluding remarks will include limitations of the study and recommendations for further research.

2 Literature Review

2.1 Overview

This chapter provides a literature review of vocabulary knowledge and a literature review of previous research on intentional and incidental vocabulary learning, the *forgetting curve*, and rehearsal techniques, and a literature review of techniques for supporting vocabulary learning. By the end of this chapter justification for the study is made, a summary of key issues is given, and research questions and predictions of the results are provided.

Learning a good proportion of vocabulary items is important for effective communication and understanding and is crucial for language use (Laufer, 1996; Nation, 2011; Schmitt, 2010). Vocabulary knowledge is a fundamental component of second language general competence and it is essential for successful communication (Nation, 2001). Although the use of communicative language teaching has become more popular than intentional vocabulary learning, i.e., "vocabulary memorization", empirical studies demonstrate that techniques of learning vocabulary intentionally are more efficient (Elgort, 2011; Hulstijn, 2001; Nakata, 2008). The following is a review of literature concerned with vocabulary knowledge.

2.2 Vocabulary Knowledge

Lexis is the Greek synonym of word and it "refers to all the words in a language, the entire vocabulary of a language" (Barcroft, Schmitt, & Sunderman, 2011, p. 571). *Types, tokens, lemmas, word families,* and *hapax legomena* are also terms used to describe words (Milton, 2009), and are discussed in section 2.2.3.1. It is essential to know what type of word knowledge the researcher is seeking from learners and what kind of words are measured as this will determine the composition of learners' word knowledge. Firstly, *receptive* or *passive* knowledge and *productive* or *active* knowledge (Milton, 2009) explained in section 2.2.1. The second common convention of vocabulary knowledge is *breadth* and *depth* (Anderson & Freebody, 1981) which is explained in section 2.2.2. Following this *word families* and other word terms are addressed.

2.2.1 *Receptive* and *productive* vocabulary knowledge

Receptive or *passive* knowledge is the knowledge of words when listening or reading a language, while *productive*, or *active*, knowledge involves recall (i.e., to bring retained information back from memory) of words when speaking or writing (Daller et al., 2007; Milton, 2009; Nation, 2013). Schmitt (2010, p. 87) considers *meaning recall* and *meaning recognition* to be *receptive* vocabulary knowledge, and *form recall* and *form recognition* to be *productive* vocabulary knowledge. Nation (2013) provides a list of aspects of what is involved in knowing a word and distinguishes between these aspects in *receptive* and *productive* knowledge; he also divides general knowledge into form, meaning, and use (see Table 2.1).

Table 2.1 *What is involved in knowing a word? (from Nation, 2013, p 49)*

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 recognizable in this word?
		P	What word 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 other 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 pattern does the word occur?
		P	In what pattern must we use this word?
	collocations	R	What words or types of words occur with this one?

Table 2.1 *What is involved in knowing a word? (from Nation, 2013, p 49)*

		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?

R = receptive P = productive

2.2.2 *Breadth* and *depth* of vocabulary knowledge

Breadth of vocabulary knowledge refers to the number of vocabulary items known (Milton, 2009). However, the *depth* of vocabulary knowledge involves the knowledge of many aspects of a word (Daller et al., 2007; Milton, 2009; Nation, 2013), such as concepts and referents, associations, grammatical functions, collocations, and constraints on use (Daller et al., 2007). The definition of knowing a word is simple, passive, word recognition (Daller et al., 2007; Milton, 2009). According to Scarborough and Brady (2002), word recognition is "a somewhat broader term that refers to the process of determining a written word's identity by any means, not just by using one's knowledge of orthographic regularities". Daller et al. (2007) and Milton (2009) state that being able to recognize the form of a word and distinguishing it from non-words (i.e., a group of meaningless symbols) is the basic type of word knowledge. If a learner is able to recognize and distinguish a word from a non-word, s/he is at the first step of word knowledge and more complex word knowledge can follow.

Certainly, word knowledge is closely related to overall language proficiency. In order to have adequate language comprehension, and to benefit fully from learning vocabulary, the number of words learned needs to be standardized (Nation, 2013). Learners need to cross a vocabulary threshold to have successful reading comprehension (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2013). The minimum vocabulary threshold is 95%, while the optimal is 98% coverage of a text (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2013). This means that to achieve the 95%, a learners must know 3000 word families for spoken texts, and 4000 word families for written texts, and to achieve 98% coverage learners need to know 6000-7000 word families for spoken texts, and 8000-10000 word families for written texts (Nation, 2013) (see Table 2.2).

Table 2.2 *Vocabulary sizes needed to get 95% and 98% coverage (including proper nouns) of various kinds of texts (from Nation, 2013, p 208).*

texts	95% coverage	98% coverage	proper nouns
novels	4,000 word families	9,000 word families	1-2%
newspapers	4,000 word families	8,000 word families	5-6%
writing for children	4,000 word families	10,000 word families	3.5%
children's movies	4,000 word families	6,000 word families	1.5%
spoken English	3,000 word families	7,000 word families	1.3%

2.2.3 Characteristics of vocabulary items

2.2.3.1 Word families

In addition to the researcher's selection of the type of word knowledge s/he is looking for in a research study, it is essential to be aware of the type of word unit measured in an experiment (Daller et al., 2007; Milton, 2009). Being aware of the word type is important in order to make precise decisions about the vocabulary material and the vocabulary knowledge tests (Milton, 2009). It is not easy to define a word, however, there are many attempts in the literature to define and classify words in order to use them as units of measurements (Milton, 2009). For example, *tokens* (i.e. running word) and *types* are used to count words in a text, they provide information about the length of a passage, an essay, or a corpus, and to give a measurement of the size of the corpus (Daller et al., 2007; Milton, 2009). The difference between *tokens* and *types* is that *tokens* are used to measure the occurrences of every word in a text, but *types* are units used to count the number of words but not the number of their occurrences. *Hapax legomena* are rare words that occur only once in a corpus (Milton, 2009).

However, if the goal is to measure learners vocabulary knowledge, for example, it is useless to count words that are repeated, or ones that are for grammatical usage, such as *the* (Milton, 2009). Therefore, researchers such as Bauer and Nation (1993), Coxhead (2000), Diack (1975), Goulden, Nation, and Read (1990), and Seashore and Eckerson (1940), created units of measurement to serve purposes other than for measuring the length of a text and to produce smaller, more manageable, figures for counting. For example, Bauer and Nation (1993) provide a list of nine bands of affixes to differentiate between *lemmas* and *word families*, and

Coxhead (2000) created a word list for non-native speakers applying a *word family* as a word unit.

There is a debate about the preference of the use of *lemmas* or *word families* as word units in vocabulary tests. According to Nation and Waring (1997 p.9) "a *lemma* is a base word and its inflected forms". The inflected forms of a *lemma* are limited, for example the *lemma to work* is a baseword and *working*, *works,* and *worked* are its inflected forms (Daller et al., 2007). *Lemmas* include the baseword and its inflected forms without changing its part of speech (Milton, 2009). However, *word families* are more inclusive than *lemmas* (Daller et al., 2007), and word groups consist of wider groups of derivations and inflections (Milton, 2009). A *word family* is defined as "the base form of a word plus its inflected forms (third person -s, -ed, -ing, plural -s, possessive -s, comparative -er and superlative -est) plus derived forms made from certain uses of the following affixes (-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-, -al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, in-)" (Hirsh & Nation, 1992, p.692). Therefore, the *lemma* of the word *govern* includes *governs*, *governing,* and *governed,* and the *word family* includes the same inflected forms plus others, such as *government* and *governor* (Treffers-Daller, Parslow, & Williams, 2016).

Diack (1975) and Goulden et al. (1990) replace the use of *lemma*, as a word unit, with *word family* to make it easier to measure native speakers vocabulary knowledge. Nation uses *word family* as a word unit to measure vocabulary size in the *Vocabulary Size Test* (VST) because he claims that *lemma* is a more appropriate word unit to measure productive vocabulary knowledge, and *word family* is more appropriate to measure receptive vocabulary knowledge (Nation & Beglar, 2007). Nation assumes that if the learner knows the noun *walk* they will be able to discover the verb *walk,* and the adjectives *walking* and *walker,* which are all part of the same *word family*, but not parts of the same *lemma* (Nation, 2012). Nagy, Anderson, Schommer, Scott, and Stallman (1989), Bertram, Baayen, and Schreuder, (2000), and Bertram, Laine, and Virkkala (2000) suggest that *word families* are psychologically real (as cited in Nation & Beglar, 2007). Another reason is that vocabulary learners with limited language proficiency can control word-building processes and, therefore, know the meaning of the base word and its inflectional and derivational forms (Nation, 2012).

In contrast, to measure breadth of vocabulary knowledge for lower level learners, *lemmas* are used widely in check list tests as a word unit (Milton, 2009). *Lemmas* are the most reliable word unit to measure elementary and intermediate levels of language proficiency (Milton, 2009; Vermeer, 2004). Learners in early stages learn regular affixes (e.g., -s, -ing, -er) in *lemmas* and would be unable to use rare affixes (e.g., -ation, -ess, -ful) used in *word families* (Milton, 2009). Furthermore, if a learner knows a member of the word family it does not mean that s/he knows all members of the word family (Treffers-Daller et al., 2016; Vermeer, 2004).

2.2.3.2 Word frequency

Although word frequency is clearly significant, it is a characteristic of words that has only recently become a focus for many vocabulary researchers. Word frequency is important because it helps to determine what words learners need to learn and it can be a tool used to measure vocabulary knowledge (Milton, 2009; Nation, 2011). Furthermore, frequency has a key influence on performance in Lexical Decision Tasks (LDT) and naming tasks (Harley, 2008). Word frequency is central to the design of Nation and Beglar's (2007) VST. The VST measures knowledge of 14,000 English *word families* based on 14 frequency levels in the British National Corpus (BNC); it identifies the frequency levels with which the participants are unfamiliar, in addition to measuring the estimated overall vocabulary size.

Teachers cannot teach mid-frequency vocabulary (i.e., words from 4000 to 8000 word frequency level according to the BNC) to learners as it is impossible to explicitly teach a great number of lexical items (Pellicer-Sánchez & Schmitt, 2010; Saragi, Nation, & Meister, 1978). This can also apply to low-frequency words (i.e., 8000 to 14000). Therefore, learners should know how to learn independently, without the help of a teacher, when they reach the mid-frequency level because everyone has their own capabilities and have different encounters with words in textbooks (Saragi et al., 1978). Scanning a page of a book, or scanning a corpus, shows how words such as *the* and *a/an* are more frequent than words such as *curiosity* and *gravel* (Milton, 2009). Learners can easily become familiar with the most frequent words, and the opposite with low frequency words (Milton, 2009). It is important to consider that the theme of a text plays an important role in the word frequency as frequency may change in different texts on different subjects and cannot be completely equated with difficulty (Milton, 2009).

In general, higher frequency words are easier to recognise, and it is more likely that lower frequency words are harder to recognise (Harley, 2008). A corpus frequency of a word is an estimation of the average person's use of a word and it positively and strongly correlates to the subjective estimation of the frequency of the use of the word (Cleland, Gaskell, Quinlan, & Tamminen, 2006). It is important to note that not all high frequency words are easy to recognise and not all low frequency words are difficult to recognise, because some high frequency words are highly polysemous (De Cock & Granger, 2004). While the learning burden (Nation, 2013) of a word depends in part on its frequency of occurrence, this is not the only factor that plays a role (see also section 2.2.3.3). Word recognition is affected by many features at the same time such as word length (Field, 2003), resemblance to the first language (L1), whether it is concrete or abstract (Laufer, 1996), and what is called the *list effect*, as learners might remember some words better than others because of their position in a list (Nakata, 2008).

Word frequency of spoken language is different from word frequency of written

language, and lists are approximate as people use their vocabulary in different life experiences (Harley, 2008). For example, it may seem that short low frequency words are easier to learn, especially they do not require an effort in listening, reading or articulating because of their shortness, but in fact, most low frequency words are difficult because of the lack of exposure to them. (Milton, 2009). Words pronunciation, its resemblance to the first language (L1), whether it is concrete or abstract affect the ease or difficulty of learning depending on the groups of learners circumstances but frequency has a general impact on learning (Milton, 2009).

According to Nation (2001) eighty per cent of running words (i.e. *tokens*) are high frequency words, therefore learning high frequency words will help learners achieve knowledge of a large per cent of running text. However, it is often important to learn word lists for specific purposes, for example, Nation (2001) states that five percent of the running words in general texts are technical words used for specific purposes and it is useful for specialists in a specific area to learn these technical words. Furthermore, he states that academic words found in textbooks are found in nine percent of the running words in general texts and it is effective for learners of a L2 in academic contexts to learn this nine per cent of words in the L2. Nation (2001) recommends the 570 highly frequent academic words in academic contexts created by Coxhead (2000) for English as a second language (ESL) learners. Teachers should not spend long periods of time explicitly teaching words with moderate frequency and should focus on the learners' needs. For example, teachers might help learners learn proper names, general low frequency words that may be formal, old fashioned and rare dialect words, and technical vocabulary in which they may be interested or are in their area of specialisation (Nation, 2001). Teachers can select and teach explicitly a limited number of vocabulary items they believe the learners need in terms of their subject area, proficiency or interest and can guide the learners of ways to learn a wider range of vocabulary items independently without their help.

2.2.3.3 Difficulties for Arab language learners of English

ESL learners in general have a number of difficulties, such as word length, words with low frequency derivational morphemes, and the learning burden. Longer written words require more time to be memorised and to be phonologically recoded (Field, 2003). Furthermore, the words that have suffixes with high frequency are more likely to be learned than words that have suffixes with low frequency (Nation, 2001). Not all words are equally difficult to process because the learning burden is different for each word. Nation (2013) defines the learning burden as "the amount of effort required to learn a word" (p. 10). The learning burden of a word depends on many factors such as, the language proficiency of the learner, the amount of time

allowed to learn a number of words, the word's resemblance to the learners L1 and the frequency of exposure to learn a word. Apparently, the notion of the learning burden was first introduced and defined by Swenson and West (1934) (as cited in Nation, 2013). Nation and Webb (2017) state that measuring vocabulary knowledge means measuring the difficulty of the newly learned word. They claim that frequency of repetition is one aspect of that which affects the learning burden.

Furthermore, differences between L1 and L2 can facilitate or hinder learning. The similarity between English and Arabic languages can be seen in the characteristics of the writing system in which graphemes stand for phonemes (Saigh & Schmitt, 2012). However, the writing system itself is different in Arabic from English. The writing system in Arabic starts from the right side and ends on the left side, while it is the opposite in English. Unlike English, capital letters and italics do not exist in Arabic. Furthermore, in Arabic letters have a cursive connection feature which is obligatory for most letters (Catford, Palmer, McCarus, Moray, & Sinder, 1974). For example, in one word most letters connect to the previous and the following letters. Unlike English, printed, typed, or handwritten Arabic have similar forms for the letters. Arabic has the distinctive feature where letters look the same but are distinguished by the placement of dots (e.g., ب/b/, ت/t/, ث/θ/). Also, most cursive connections change into four shapes (e.g., ب, بـ , ـب, ـبـ) according to the letters before or after, and according to the position of the letter in the word (Catford, Palmer, McCarus, Moray, & Sinder, 1974).

Bassetti and Cook (2005) describe English as *alphabetic* and Arabic as *consonantal* (p. 5). The Arabic writing system is relatively transparent as the phoneme (i.e., sound) represents the grapheme (i.e., symbol) most of the time. English phoneme-grapheme relationship is less transparent as sounds are not always represented by symbols (Saigh & Schmitt, 2012). From the 26 letters in the English alphabet, 21 are used to represent 24 consonants and only five are used to represent 11 vowels, not including the diphthongs (Carney, 2012). However, the 28 letters in Modern Standard Arabic represent all 28 consonants. The Arabic writing system has eight vowels (three long, three short, and two diphthongs) (Swan & Smith, Bernard, 2001). The three long vowels are represented by letters, but short vowels are represented by diacritical marks (Abu–Rabia, 2002). English vowels are always written and contextual knowledge facilitates knowing vowels for Arab learners if they are missing: (e.g., nglsh spkrs cn ndrstnd wrds wrttn ntrly wth vwls, spclly whn thy cm n sm knd f cntxt). However, using contextual knowledge is not easy for ESL learners and this can interrupt their fluency when reading (Saigh & Schmitt, 2012). According to Swan, Smith and Bernard (2001), the most common error Arab ESL learners make is to replace the sound /I/ with the sound /e/, e.g., in the words *bit* and *bet.* Also, Arab ESL learners confuse the sound /ɒ/ with the sound /ɔː/, e.g., in the words *cot* and *caught.* Furthermore, Arab ESL learners pronounce the diphthongs /eɪ/ and /əʊ/ in a short

manner (/e/ and /ɒ/), e.g., in the words *raid* and *hope*, making them sound like *red* and *hop* (Swan & Smith, Bernard, 2001).

Arab ESL learners have more difficulties in the reading and literacy skills of English, but they find listening and speaking easier in comparison to other ESL learners (Fender, 2008). Arab ESL learners struggle with English spelling in particular and this affects their reading (Fender, 2008) and writing skills (Saigh & Schmitt, 2012). Fender (2008) states that there is a relationship between phonological and orthographic decoding of words and word recognition. The Arab ESL learners ability to decode phonological and orthographic aspects of English words affects their ESL ability to read, to acquire vocabulary, and to pronounce words proficiently, therefore these skills affect word recognition (Fender, 2008). Arabic words can be read without diacritical marks as readers can use the morphological knowledge, syntactic knowledge, and sentence and discourse contexts to process its phonological features (Abu–Rabia, 2002). This skill is transferred to L2 by Arab ESL learners, but causes difficulty in the recognition of the English vowels and results in errors because diacritical marks do not exist in the English language (Abu–Rabia, 2002).

Word-level processing refers to the component of sentence processing and involves the holistic processing of a word (Allen, Stadtlander, Groth, Pickle, & Madden, 2000). L2 word-level processing development is different from L1 word-level processing development. The development of L1 word-level processing affects ESL word-level processing skills and therefore leaners differ according to the level of their L1 skills (Fender, 2003). Saigh and Schmitt (2012) state that spelling (i.e., phonological and orthographic decoding) are also important for both *receptive* and *productive* word knowledge. Arab ESL learners struggle to learn a word form for several reasons: they often transfer their L1 (i.e., Arabic) knowledge to the L2 (i.e., English) when trying to interpret the L2 word forms. This can be successful sometimes, but can also cause errors in learning the second language (Saigh & Schmitt, 2012).

Arab ESL learners show particular difficulty in recognizing English vowels in a phenomenon called *vowel blindness*. In *vowel blindness* Arab ESL learners do not know the correct vowel to use and this cause difficulties in word recognition (Saigh & Schmitt, 2012). In a study conducted by Saigh and Schmitt (2012) they investigate learners transfer of their L1 knowledge of vowels to L2 spelling and how this affects the way they recognize words. They investigate Arab ESL learners' reactions to short and long English vowels and find that when Arab ESL learners are introduced to English words with missing vowels they can identify the spelling errors (e.g., cptain), but when a vowel is replaced with a wrong vowel they find it more difficult to identify the error (e.g. exclasively). Saigh and Schmitt's, (2012) results also show that Arab ESL learners have more difficulty with short vowels than with long ones.

Not being able to identify the correct vowel is a common error for ESL learners in general due to the similarity of word forms (Laufer, 1988). Laufer (1988) investigated a phenomenon called 'synforms' and found that some similar word forms are confusing for ESL learners, particularly similar word forms which have different vowels (e.g., *ingenuous* and *ingenious*, *affect* and *effect, expert* and *excerpt*) or similar word forms with different suffixes (e.g., *economic* and *economical, industrious* and *industrial*).

In a study conducted by Fender (2003), a group of native Arab ESL learners were successful in comprehending and using words in long sentences, but they were significantly slower in recognizing words in the LDT than a group of Japanese ESL learners. The results of this study indicate that Arab ESL learners may have problems with English orthography as it is inconsistent. In other words, many English words have irregular spellings, unlike Arabic (Fender, 2003). Orthographic neighbourhood and phonological similarity is another problem facing Arab ESL learners (Saigh & Schmitt, 2012). Unfortunately, beginner Arab ESL learners struggle with word recognition because they depend on interpreting every phonological aspect of a word and this struggle affects the ease of lexical access (Saigh & Schmitt, 2012). Words with a large orthographic neighbourhood create more confusion, or even short words that have a similar word form but a different meaning (Heuven, Dijkstra, & Grainger, 1998). For example, it can be hard for Arab ESL learners to differentiate between the words *tape, tale, tame, lake, sake, tack* (Heuven, Dijkstra, & Grainger, 1998). However, orthographical and phonological patterns in the L1 that are similar to orthographical and phonological patterns in the L2 are easier to learn (de Groot, 2006). For example, it might be easier for Arab ESL learners to recognise the words *mum, lamp, tomato, sugar* as their phonological features are similar to the Arabic words /mama/, /lamba/, /tamatəm/ and /sokar/.

Arab ESL learners transfer their phonological knowledge of the L1 to the L2 and they transform the way they write the sounds from L1 to L2 spelling and this can improve or hinder learning the L2 (Saigh & Schmitt, 2012). The more the phonological/orthographical systems of the L1 and L2 are similar the more positive the transfer will be, and the more the phonological/orthographical systems of L1 and L2 are different the more negative the transfer will be. This problem may be resolved when the learner is exposed to the English language and its phonological and orthographical system more frequently. Arab ESL learners transfer L1 phonological systems to a greater extent in comparison to other ESL learners (Fender, 2003). Randall and Meara (1988) conducted three studies in which they investigated the reaction of Arab learners of Roman letters and found that Arab ESL learners use their knowledge of Arabic sounds and letters to process Roman letters.

Another difficulty with which Arab ESL learners struggle concerns initial consonant clusters in English. A consonant cluster is a syllable in a word with two or more consonants

occurring successively (Jayaraman, 2010). Every language allows specific combinations of phonemes. English, for example, has its own sequences of consonant clusters which may consist of two, three, four or even five consonants occurring together (Fromkin, Rodman, & Hyams, 2010). However, in Arabic, three or more consonants in a cluster does not exist. Furthermore, clusters never occur in initial syllable position in a word. Arab ESL learners find the cccv words and cccvcc words, where the consonant cluster is in the initial, medial, or final position, very difficult (c stands for consonant and v stands for vowel) (Jayaraman, 2010).

Arab ESL learners avoid the difficulty of pronouncing initial consonant clusters in English by inserting an epenthetic vowel between those consonants in the initial cluster (Catford et al., 1974). It is difficult for Arabic speakers, especially at the beginner level of learning English, to pronounce consonants clusters because they do not exist in their L1. This L1 transfer causes errors such as pronouncing the word *skate* as /sIket/, the word *gleam* as /gəlim/, the word *true* as /təru/ and the word *express* as /ɛksəprɛs/. Moreover, an initial three consonant cluster can create another problem where Arab ESL learners might insert a vowel between the first two consonants and make the last consonant a new syllable (e.g., *strike* might be pronounced /sɪtraɪk/) (Catford et al., 1974). However, Saiegh-Haddad (2003) points out that some colloquial Arabic permits an initial consonant cluster, such as in the word /tra:b/ (*soil*) but he points out that some Arab find it difficult and pronounce it as /tura:b/. Furthermore, he points out that colloquial Arabic speakers insert vowels in Arabic words that consist of a final two consonant cluster. For example, the word /sahl/, meaning *plain,* becomes /sahel/ and the word /qalb/, meaning *heart,* becomes /qaleb/.

Another difficulty for Arab ESL learners is sounds in English that do not exist in Arabic. Learners tend to create phonetic sounds when they want to use phonetic sounds that do not exist in the phoneme system of the Mother Tongue. Catford et al. (1974) give examples of how Arab ESL learners make errors by creating phonetic sounds close to those found in their L1 when pronouncing words in the L2. For example, the English consonant sound /p/ is a voiceless stop and is a different phoneme from the English sound /b/ which is the voiced stop. The phoneme /p/ does not exist in Arabic, it only occurs as an allophonic variant of the sound /b/. Arab ESL learners also might use the phoneme /f/ instead of the phoneme /v/ as the latter does not exist in the Arabic language. Swan and Smith (2001) also state that the sound /g/ might cause confusion with the sound /k/ and this confusion can potentially cause difficulty in learning, and the sound /r/ in Arabic is different from the voiced flap /r/ in English. Arab speakers tend to over pronounce the post-vocalic /r/, such as in the words *car* and *park*. This study focuses on the performance of Arab ESL learners and therefore their specific difficulty they have in learning English are considered.

2.2.4 Measuring vocabulary knowledge

Researchers as well as practitioners are interested in measuring vocabulary size for native and non-native speakers because a large vocabulary size is often seen as a proxy for general language proficiency (Meara, 1996). There are several vocabulary size tests that are used in vocabulary research but not all these tests are validated. Designing and improving vocabulary tests are important for teachers and researchers (Read, 2000). Vocabulary test designers must explicitly justify the purpose of the test and show the extent to which the design and purpose of the test is related by analysing and committing to vocabulary test validation frameworks (Read & Chapelle, 2001). Test validation can be achieved by committing to three components: first is the construct definition, second is the performance summary and reporting, and third is test presentation (Read & Chapelle, 2001).

Read and Chapelle (2001) explain that the validity of a vocabulary test relies on defining its construct because the interpretation of the results is strongly related to the test construct. Chapelle (1998) draws on Messick's (1981, 1989) work to highlight three ways of defining the construct of a vocabulary test: a *trait definition, a behaviourist definition* and an *interactionalist approach definition* (as cited in Read & Chapelle, 2001).

Under a *trait definition* of vocabulary, this component of language proficiency is seen as a specific trait of a learner. Contextual variables (that is differences in the purposes for which vocabulary is used) are not of central importance in this view of the construct. In contrast, in the *behaviourist approach* to the construct, the context plays a significant role in that the test designer attempts to predict the learner's future performance by replicating real context in which the learner needs to use language (e.g. as a nurse in a hospital or as a teacher in a classroom). The reason for replicating real contexts in tests with behaviourists construct is because test designers think that learners' underlying knowledge is too difficult to define and it is easier to define the context in which the learner might use their hidden knowledge of language. Therefore, in this approach, vocabulary is not highlighted or even explicitly mentioned as a separate skill. In the third approach to the definition of the construct, the *interactionalist definition,* the test designer considers both trait and context: vocabulary is defined as an underlying trait, but at the same time it is related to a particular contexts in which it is used. According to Read and Chapelle (2001) the *interactionalist approach* should be adopted more widely as learning contexts differ so much from each other that an approach which sees vocabulary simply as a discrete trait (irrespective of the context) is no longer appropriate.

The second component of test validation refers to the purposes for which the tests are used (e.g. practical outcomes of tests and the uses of the test results).The third component of test validation is the test presentation or the impacts of the test. It is important to consider the way, time, and the consequences of testing and who are the test-takers (Milton, 2009; Nation, 2012; Read, 2007; Schmitt, 2010). The test design is affected by audiences of the test (e.g. program administrators, L2 teachers or researchers). For example, program administrators of an institution need to present and interpret data that is comparative to test data from other institutions (Read & Chapelle, 2001). Moreover, tests should be presented and interpreted in a way that learners, teachers, and researchers can focus more on areas that needs improvement (Milton, 2009; Nation, 2012; Read, 2007; Schmitt, 2010).

Read (2000) proposed three continua/dimensions that test designers should review when designing vocabulary knowledge measures. The first continuum (*discrete – embedded*) relates to whether the test is labelled as 'a vocabulary test' and vocabulary is specifically measured as a separate skill. In other tests, vocabulary is *embedded* as one element of the test. In such a test, the focus is not only on vocabulary knowledge but also on other language skills (Read, 2000).

The second continuum (*selective – comprehensive*) relates to whether specific vocabulary items are selected for inclusion in a test. In *comprehensive* tests all the learners' output in a task is analysed (Read, 2000). The third continuum (*context-independent – context-dependent*) relates to whether or not words are presented in isolation from linguistic context or embedded in it (Read, 2000).

For example, the Lexical Frequency Profile (LFP) by Laufer and Nation (1995) is based on a *trait definition* construct and it is a test designed to measure the accurate use of vocabulary in written compositions (Read & Chapelle, 2001). It is *discrete* as it is clearly a vocabulary test and it has a *comprehensive* construct in which all the learners' output is analysed and incorrectly used content words by the learner are excluded from the frequency analysis (Read & Chapelle, 2001). Therefore, LFP is a *context dependent* measure as vocabulary knowledge is measured in the context of a writing task.

An example of a test based on an *interactionalist definition* of the construct is the ESL Composition Profile by Jacobs, Zingraf, Wormuth, Hartfiel, and Hughey (1981) (as cited in Read & Chapelle, 2001). It is a writing proficiency test in English and vocabulary use is one of the constructs measured. Vocabulary is *embedded* in the test as the authors do not state explicitly that it is a vocabulary test. It is a *comprehensive* test as all the vocabulary in the written performance of the learners is analysed. It is rated in connection to other four scales to form an overall profile of the learners' writing performance. ESL Composition Profile has the *context dependence* feature as vocabulary knowledge is measured in context (Read & Chapelle, 2001).

According to Read and Chapelle (2001) the paper-based Test of English as a Foreign Language (TOEFL) by the Educational Testing Service, (1995) is also based on an *interactionalist definition* of the construct because vocabulary is a hidden trait and it is analysed in the specific context of academic texts. In this test, the test-taker is assessed on vocabulary items in a reading passage. Although this type of assessment might be regarded as contextualized but the test-takers might treat vocabulary items in isolation of the passage and they might not use contextual information to show their vocabulary knowledge (Read, 2000). Therefore, the TOEFL has *embedded* as well as *selective* features as the test does not measure vocabulary knowledge by itself and vocabulary items are analysed in the context of a reading task (Read & Chapelle, 2001).

Other examples of tests that are widely used are the *Vocabulary Levels Test* (VLT), the VST, *Checklist Tests*, *Peabody Picture Vocabulary Test* (PPVT), *X_Lex* test and *Y_Lext* test. The VLT by Nation (1990) is based on a *trait definition* of the construct because a specific trait (i.e. vocabulary size) of a learner is assessed and the linguistic context is not so important in its design (Read & Chapelle, 2001). It is a *discrete* test as it is labelled 'a vocabulary test' and it is *selective* because high-frequency target vocabulary items are selected to provide an estimation of the size of a learner's vocabulary. The VLT is a *context independent* test in that the test presents words in isolation of linguistic context.

The VLT uses a form–recognition matching format and it provides profile knowledge at four frequency levels: 2,000, 3,000, 5,000 and 10,000. Each level represents key knowledge of the English language (Nation, 1990). According to Schonell, Meddleton and Shaw (1956) knowledge of the 2,000 word frequency level is important for daily conversation, the 3,000 word frequency level is important for early stages in reading comprehension, and the 5,000 word frequency level is important for advanced reading. This final level represents the top of the higher frequency words relating to reading ability, while the 10,000 word frequency level represents a wider knowledge of the English language (as cited in Schmitt, 2010). According to Schmitt (2010), the VLT is a good test for placement and diagnostic purposes but not to produce an overall vocabulary size of the learner.

For the purpose of measuring learners' overall vocabulary sizes the Vocabulary Size Test (VST), developed by Nation and Beglar (2007) is frequently used. It is based on a trait definition of vocabulary as it measures decontextualised knowledge of the word by presenting it in a non-defining context word in the test (Nation, 2012).The VST is different from the VLT in that it aims to test overall vocabulary size and it also provides a good indication of vocabulary growth after learning. Using Read and Chapelle's (2001) framework, the VST focuses on *discrete, selective, relatively context-independent vocabulary* (Nation, 2012). Nation and Beglar's (2007) VST uses 14 different English frequency levels based on *word families*

occurring in the BNC; it provides information about the frequency levels with which the participants are unfamiliar, in addition to measuring the estimated overall vocabulary size.

Schmitt (2014) urges researchers not to misuse the VST and he argues that the VST does not measure the learners' ability to use the target words in reading, writing, listening, and speaking. He claims that the VST does not measure the depth of vocabulary knowledge as selecting *tests from Meara and Colleagues* is a Yes/No test where test-takers are asked to tick (√) the words they think they know and put a cross (x) against the words they think they do not know (Schmitt, 2010). This test has the *trait definition* construct (Read, 2007). The stimuli are words and non-words and they are displayed with no supporting context (Schmitt, 2010). Despite the simplicity of conducting this test and the higher sample rate it can provide, there are several limitations such as there is a high chance that the test takers might over-evaluate their knowledge of the words and write (√) for real words that they do not really know (Schmitt, 2010).

X_Lex is a checklist test that contains words from the first 5,000 word frequency levels (Milton, 2009; Nation, 2012; Read, 2007; Schmitt, 2010). It tests the knowledge of 120 vocabulary items, 100 of which are real words and 20 are non-words (Schmitt, 2010). The test-takers respond to the words they know by clicking *yes*. This test was criticized in that test-takers might over-estimate their knowledge of words and click *yes* while they do not really know the word's meaning. However, *X_Lex* is beneficial in highlighting the test-takers knowledge of certain vocabulary levels, and by assisting the process of designing additional tests targeted at a level appropriate for those learners (Schmitt, 2010). The *Y_Lex* is an improved version of the *X_Lex* in that it measures vocabulary size for advanced learners and it covers vocabulary bands from the 6,000 to 10,000 word frequency level (Schmitt, 2010).

Peabody Picture Vocabulary Test PPVT is another L2 vocabulary size measure (Schmitt, 2010). It is a meaning recognition test where test-takers (i.e. children and adults) point to a picture that represents the word provided by the experimenter (Schmitt, 2010). The test becomes progressively harder as the frequency level of target words decreases (Schmitt, 2010).

The need to utilize different vocabulary measures is important because there are different dimensions and degrees of knowing a word (Hulstijn, 2010; Milton, 2009; Nation, 2001; Schmitt, 2010). One of the most salient differences of measuring word knowledge is measuring *receptive* and *productive* vocabulary knowledge. Researchers often differentiate between a learner's "passive or receptive" vocabulary knowledge, seen when learners recognize words while listening or reading, and their "active or productive" vocabulary knowledge, seen when learners speak or write (Milton, 2009; Nation, 2001).

Meara (1990) suggests that there is no clear cut difference between the *depth* and *breadth* of vocabulary knowledge as they can meet in the middle of the continuum of word knowledge. For example, if a learner who knows the word *sofa* and *chair* also learns *bench*, they obtain a deeper knowledge of what *sofa* means. The learner extends the number of words learnt in one semantic domain by restructuring the knowledge obtained in that domain. Therefore, measuring *depth* and *breadth* requires caution as they overlap.

One of the vocabulary tests that was designed to measure *productive* vocabulary knowledge at different frequency is the Productive Vocabulary Levels Test (PLT) (Laufer & Nation, 1999). The LDT can be used to measure *receptiv*e knowledge of the form of the word. The LDT is "one of the most established paradigms for studying processes involved in word recognition and the structure of the mental lexicon" (Elgort, 2011, p. 375). The process of repeating word form and meaning improves retrieval and therefore improves the accuracy and speed of word recognition (Elgort, 2011).

Vocabulary test designers use different word units for different purposes and there is no comprehensive vocabulary test that can tap into all aspects of the constructs of vocabulary (Milton, 2009). However, there are well established tests that can measure different aspects of vocabulary knowledge (Milton, 2009). Nation's VST is one of these tests which is used as a general measure of vocabulary size and is one of the most widely used (Beglar, 2010; Nation & Beglar, 2007).

2.3 Intentional and Incidental Vocabulary Learning

According to Schmidt and Lee (2005 p. 465) learning "is a set of internal processes associated with practice or experience leading to relatively permanent changes in the capability for skill". There are two learning processes that have been compared in L2 acquisition research for a long time: intentional learning and incidental learning. The following is a detailed explanation of the two types of learning followed by a comparison of both types. According to Hulstijn (2003 p.349), intentional L2 vocabulary studying involves "the deliberate committing to memory of thousands of words (their meaning, sound and spelling)". Intentional vocabulary learning involves the teacher in the design or plan of a vocabulary activity and the purposeful encouragement of learners in ways of memorizing words. It also means that the teacher points the learners attention to vocabulary learning techniques (Nation, 1990). Intentional vocabulary learning is a process in which learners deliberately memorize lists of words. Fortunately, as a result of the practicality of investigating intentional vocabulary learning, research in this area

can provide a rich source of methodologies and have pedagogical and research implications that might be able to facilitate vocabulary acquisition (Hulstijn, 2003).

On the one hand, intentional learning involves rehearsal and memorizing techniques and focuses on the form of the word rather than on the meaning it conveys in context (Hulstijn, 2003). Intentional learning happens by learning explicitly, which involves using different techniques such as word cards, mnemonic techniques, and L1 and L2 translation of words (Dekeyser, 2003; Elgort, 2011; Nation, 2011). Intentional vocabulary learning refers to activities conducted specifically for vocabulary learning (Nakata, 2008) and requires controlled consciousness (Hulstijn, 2003; Nation, 2011; Schmitt, 2000). On the other hand, incidental learning is a process of learning in which a learner wants to learn something, but eventually learns something else without the intention to do so (Richards & Schmidt, 2002). Incidental vocabulary learning has become the commonest way to learn vocabulary because of the focus on communicative language teaching, this has led to intentional learning being long underestimated and misunderstood (Hulstijn, 2001; Nakata, 2008; Nation, 2002, 2013).

Some researchers agree with Krashen (1989) that the characteristics of both intentional learning and incidental learning are somewhat distinct. Incidental learning focuses more on meaning, in particular on guessing the meaning of a word from written or spoken text (Hulstijn, 2003; Nation, 2011; Schmitt, 2000). It occurs when a learner acquires new vocabulary items from communicative activities involving comprehensive reading or listening (Elgort, 2011; Hulstijn, 2003; Nation, 2013). It involves learning by focusing on the meaning of language and storing forms unconsciously then using them automatically later (Hulstijn, 2003; Nation, 2013).

2.3.1 Evaluation of intentional and incidental learning research

From the 1980s onwards, intentional learning became a less fashionable way of learning and incidental learning became more popular (Elgort, 2011). A significant study by Nagy, Herman, and Anderson, (1985) provides evidence that being sufficiently exposed to written language can result in incidental vocabulary learning from context. Following this study many researchers claim that incidental vocabulary learning from context is more effective than intentional learning (Day et al., 1991; Jenkins, Stein, & Wysocki, 1984; Saragi et al., 1978).

A key opponent of intentional learning is Krashen (1989) according to whom linguistic knowledge is acquired when learners focus on the meaning of the language rather than the form. Krashen (2003) states, in his input hypothesis, that it is vital to receive a comprehensive input through extensive reading for better language acquisition and reading can help develop not only reading skills and fluency but also the acquisition of vocabulary and grammatical structures.

There is an ongoing debate in L2 vocabulary learning research regarding the efficiency of intentional L2 vocabulary learning. The dispute relates to whether learners should intentionally retain L2 vocabulary or rely on retaining L2 vocabulary incidentally through reading and listening practice (Hulstijn, 2003; Nation, 2011; Schmitt, 2000). A study that drives researchers' attention to incidental vocabulary learning is Nagy et al., (1985), whose results demonstrate that from a few exposures to unknown words in a written context a learner can learn new vocabulary, and that learning from written context is as effective as teaching vocabulary directly. A few years later, Krashen (1989) appeared as a key opponent of intentional learning. He claims that linguistic knowledge is acquired when learners focus on the meaning of the language rather than the form. He also notes that vocabulary acquisition occurs when learners are engaged in communicative activities that trigger unconscious acquisition (1989).

It has been seen that incidental vocabulary learning based on reading may actually inhibit vocabulary learning because learners may create false assumptions when they receive inaccurate feedback (Hulstijn, 2003). According to Nation and Coady (1988), learners may understand the overall meaning of the passage and avoid learning a new word because the meaning of it becomes less important. Guessing the meaning of new words from context can be faulty (Laufer & Sim, 1985). Furthermore, incidental learning is not suitable for learners at all proficiency levels because they need knowledge of a wide proportion of vocabulary to enable them to guess the meaning of vocabulary in context (Hulstijn, 2003; Nation, 2011, 2013). Reading is slow and haphazard (Nakata, 2008), it is time consuming (Hulstijn, 2003), and teachers can feel guilty about devoting a significant amount of time in class to extensive reading without teaching vocabulary directly (Nation, 2011). Moreover, poor training means that teachers are sometimes unable to differentiate between extensive reading, which is reading long texts for enjoyment without concentrating very much on the meaning, and intensive reading, which is the reading of shorter texts while concentrating on the meaning and details of the text, and are not fully able to teach incidental L2 vocabulary (Nation, 2011).

In comparison with incidental vocabulary learning, intentional vocabulary learning is suitable for learners at all levels of language proficiency (Nation, 2011). Nation (2013) suggests that intentional vocabulary learning is more efficient (i.e., the learner learns many words) and effective (i.e., it provides usable knowledge) in the early stages of learning than is incidental learning. He also suggests that learners need to establish a generalized, underlying concept for high frequency words before learning from context. In other words, intentional learning results in better learning than incidental learning at the elementary level (Laufer, 2003). Recent research has shown that too much has been expected from incidental learning and that retention rates of intentional vocabulary learning can result in higher vocabulary retention than incidental

vocabulary learning (Cobb, 2008; Horst et al., 2005; Hulstijn, 2003; Laufer, 2005; Nation, 2001; Schmitt, 2010). In fact, contrary to what many researchers claim, that incidental vocabulary learning from context is more effective than intentional learning (e.g. Day et al., 1991; Jenkins, Stein, & Wysocki, 1984; Saragi et al., 1978), intentional learning of words can provide rich context of new vocabulary, such as in the use of word cards and the *Leitner system* and can be more beneficial for learners than incidental learning from context (Mondria & Mondria-De Vries, 1994).

Memorizing sets of words when preparing for a new job (Elgort, 2011), or to work in a specialized area (Nation, 2001), or in a country where the exposure to the L2 is limited (Nakata, 2008), such as in Saudi Arabia for learners of English, can be achieved efficiently with intentional vocabulary learning techniques. In order for learners to know the minimum text coverage they need to commit to memory 5,000 word families (Laufer, 1989). This goal cannot be achieved effectively and rapidly without intentional learning techniques (Nation, 2011, 2013). However, in order for learners to learn only 180 words incidentally from context they probably need to read 200,000 words (Laufer, 2003).

In order to maximize the benefits of intentional learning in the classroom, teachers must be aware of how to teach vocabulary explicitly (Nation, 2013). Teachers should plan to achieve a significant improvement in their learners' vocabulary knowledge and should be aware of students' existing vocabulary knowledge when deciding on the new vocabulary lists they need to learn (Sokmen, 1997). Teachers should also give examples of words from real life contexts and make learners aware of the importance of *depth* (i.e., knowing different aspects about a word) of vocabulary knowledge (Sokmen, 1997). Another newly discovered advantage of intentional learning is that it can lead not only to explicit learning but also to implicit learning (Elgort, 2011; Nation, 2011, 2013). The former involves being aware of the learning process, while the latter does not (Dekeyser, 2003). Both explicit and implicit learning are important to enable the use the second language normally (Nation, 2011). Elgort (2011) and Nation (2013) claim that learning by using a word card method, spaced retrieval, and mnemonic techniques improves both explicit and implicit vocabulary knowledge. Finally, intentional vocabulary learning can be an addition to incidental vocabulary learning and can result in better retention rates when both types are applied together (Cohen, 1987).

2.3.2 Review of some studies of intentional and incidental vocabulary learning

In studies conducted to investigate incidental vocabulary learning, some researchers apply immediate post-test(s) to measure form and meaning recognition of target words. Studies by Waring and Takaki (2003) and Brown, Waring, and Donkaewbua (2008) measure vocabulary recognition of word form immediately after the use of graded readers by using multiple choice questions. Pellicer-Sánchez and Schmitt (2010) argue that testing the learners immediately after reading does not measure long-term learning. The results of immediate tests do not show the real amount of vocabulary gains (Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003). Pellicer-Sánchez and Schmitt (2010) state that immediate post-tests to measure incidental vocabulary learning that might occur from reading graded readers is not enough and it is crucial to apply a delayed post-test to show long-term vocabulary retention and learning. A study conducted by Waring and Takaki (2003), in which there was comparison of the results of the immediate post-test and the delayed post-test results of word recognition and word production, show recognition rate to be higher than production rate in both the immediate post-test and delayed post-test. Moreover, the recognition rate and production rate dropped in the delayed post-test, but production rate dropped sharply in the delayed post-test.

Single subject studies show successful intentional vocabulary learning. In a case study reported by Waring (2004), one subject learned 311 of 468 items after seven months of using word cards. Fitzpatrick et al. (2008) investigated spaced learning using file cards with one native English-speaking participant. The subject was asked to learn 400 high frequency Arabic words. In daily 30 minute sessions, the subject was expected to learn 20 target words over a period of 20 days until she reached 400 target words. She was allowed to revise any target word outside these learning sessions. She also recorded in a diary the techniques she used and each experience with each of the target words. Fitzpatrick et al. (2008) state that the tests were straightforward translation tasks, with the productive knowledge test administered first in all post-tests, followed by the receptive test. The subject received the tests immediately after the last learning session and then she received further tests at different times in the following ten weeks after the experiment. The subject retained 219 words out of 400 after four weeks of learning, which is very high in comparison to learning incidentally from reading, for example.

2.4 The *Forgetting Curve* and Rehearsal Techniques

It is important to deliberately plan for the time devoted to vocabulary learning (Carter & McCarthy, 1988). Teachers should play an important role in explaining learning techniques that help learners plan and be independent in learning bilingual pairs of vocabulary (Nation, 2011; Sokmen, 1997). According to Nation (1980) intentional learning using bilingual pairs can help

learners to learn 30 to 100 new words in one hour. However, these words might fade from memory if it not revised at different points in time. For long term learning and retention of those words, using word cards in the spaced repetition technique is effective (Waring, 2004).

2.4.1 The *forgetting curve*

Forgetting is basically the loss of memory (Schmidt & Lee, 2005). Forgetting is often used in research as the opposite of learning because forgetting is the permanent loss of performance or retention, but learning is the permanent gain of memory (Schmidt & Lee, 2005). A large amount of research has been conducted to investigate forgetting and memory and how these concepts are related to retention and learning. The famous theory about forgetting and retention is the *forgetting curve* investigated by Ebbinghaus in the 1880s (Finkenbinder, 1913).

In the 1880s Ebbinghaus observed and experimented his own forgetting rate (Finkenbinder, 1913). Over thirty days he devoted himself to the memorization of a set of cards with real words and non-words in different memorization sessions. He recorded his word retention at the time of each session (Finkenbinder, 1913). Thus, Ebbinghaus measured his retention of the real words and the non-words after each trial, and so took measurements after 20 minutes, one hour, nine hours, one day, two days, six days, and 31 days (Schuetze, 2015). He found that forgetting occurs after each trial, but at a different rate and he created what is called the *forgetting curve* (Schuetze, 2015). In the *forgetting curve* the retention drops sharply at the starting point of the graph followed by a progressive decline when no re-learning sessions occur (A. Baddeley, 1997; Schacter, 2002). Figure 2.1 shows how new information might fade from memory if it is not relearned.

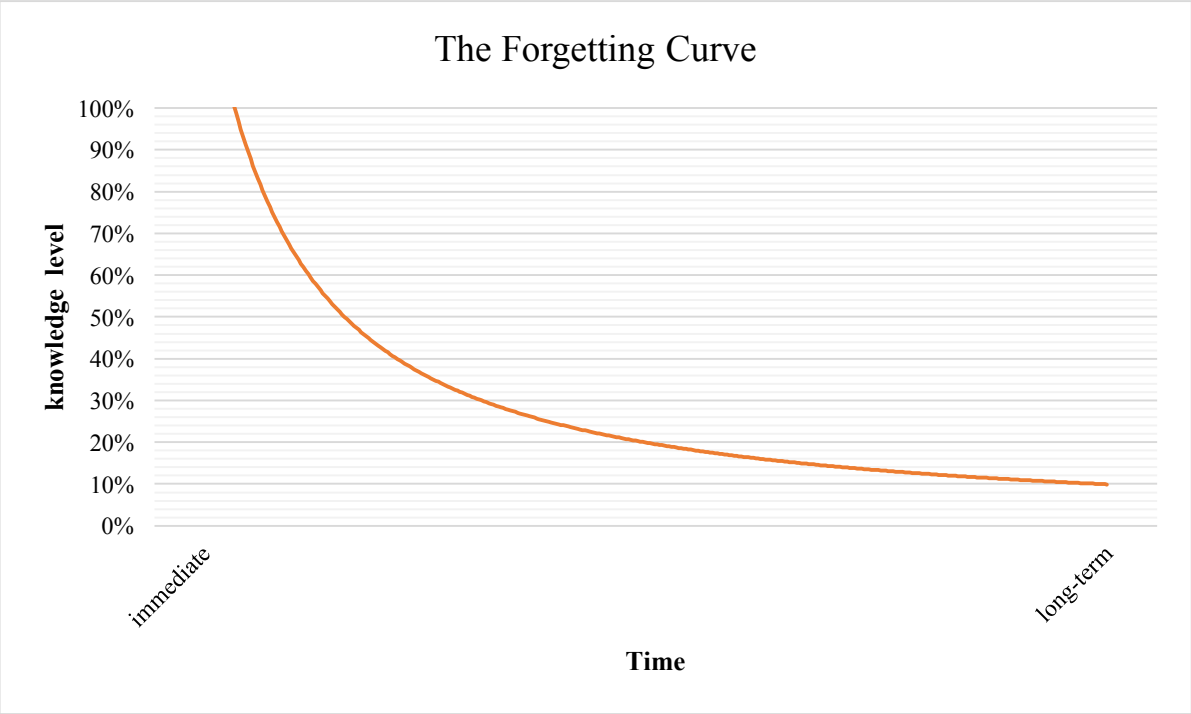


Figure 2.1 Forgetting curve (adapted from Waring, 2004)

Conversely, re-learning sessions strengthen memorization, and word retention drops mildly from one re-learning session to another. Recent learning of information will fade from memory and can be easily forgotten, but it can remain longer in the memory when repeating learning at different points of time (A. Baddeley, 1997; Pimsleur, 1967; Schacter, 2002; Waring, 2004) (see Figure 2.2).

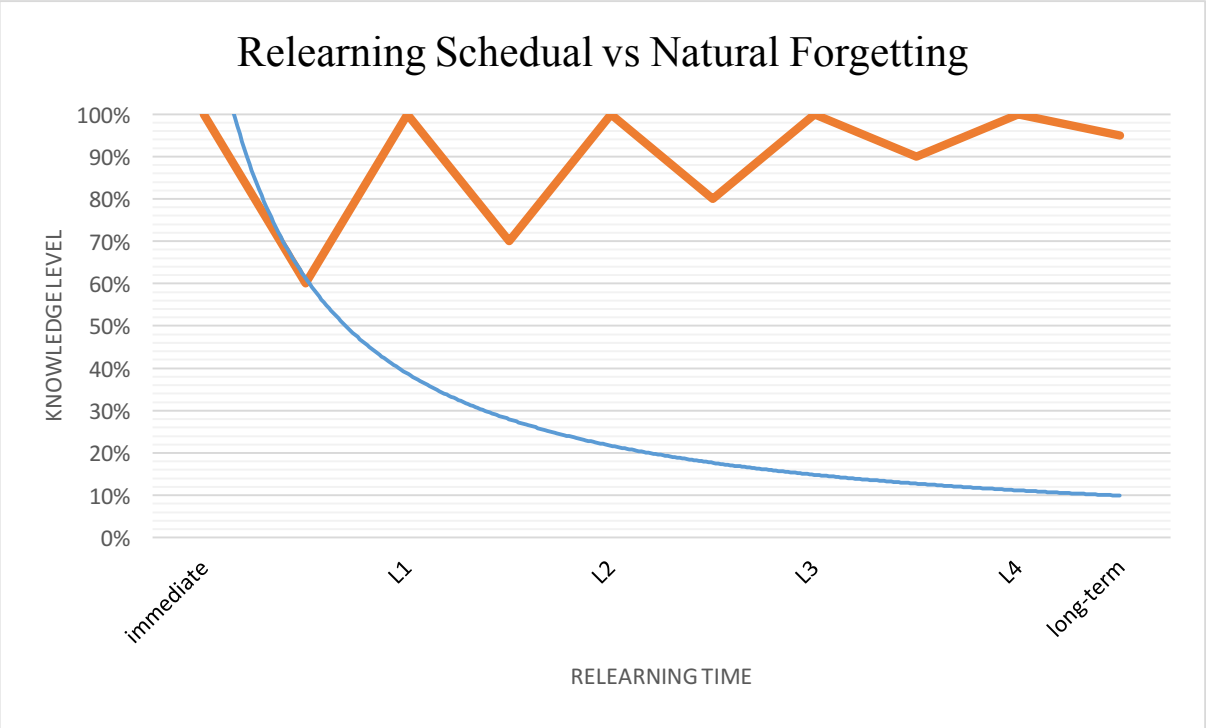


Figure 2.2 Relearning schedule vs. natural forgetting (adapted from Waring (2004))

Pimsleur (1967) investigated relearning and forgetting by being an observer and experimenter and suggests an ideal time distance between each learning session. He proposes multiplying the time between the first exposure to information and the time of recall by 5. For example, if the time of the first interval is 5 seconds, it should be multiplied by 5 and so will be reviewed after 25 seconds. Therefore, the time between each interval follows this pattern: 5x5=25 seconds, 5x25=125 seconds, 125x5= 625 seconds, and so on. The number of re-learning sessions should be sufficient and the distance between every interval should be wider until the information is retained in the long term-memory (Pimsleur, 1967).

Researchers suggest that intervals should expand by time. Lado (1964) and Oxford (1990) investigated *expanding conditions* and they suggest different timing between learning sessions. Lado (1964), in his laws of language learning, addressed the law of geometric increase of permanence. He suggests the increase by a ratio of two or three times from the time the information is first introduced to the time of correct recall of information (Lado, 1964, p. 45). Oxford (1990), in his structured reviewing plan, states the efficiency of reviewing at different intervals. He suggests re-learning after 15 minutes, one hour, three hours, one day, two days, four days, one week, two weeks, and the expanding goes on until the learner reaches a level of automaticity in learning the subject (Oxford, 1990, p. 66-67). Landauer and Bjork (1978) also conducted one of the famous experiments in expanding intervals which is explained in detail in section 2.2.2.

A study conducted by Vidal (2011) finds that repetition of reading can affect learning in comparison to other factors, such as word item, word form, and elaboration. Webb (2007b) also finds that repetition affects learning of several aspects of target words. Furthermore, Brown et al. (2008) find that adult ESL learners benefit from word repetition (between 20 to 50 times) in their vocabulary learning. However, repetition has not been given much attention because there is no consideration of other factors that can influence the validity of the repetition (Vidal, 2011). The frequently neglected factors are learner readiness, the saliency of the form, and the pressure of learning explicitly (VanPatten, Williams, & Rott, 2004, p. 15).

Repetition and rehearsal can result in the recognition of information or its recall. Recognition and recalling of information are concepts in memory research. Anderson (2015) states that it is common for recognition rates to be much higher than recall rates as recognition of words come before recall and it is not possible to recall a word without recognising it first. This can explain why taking a multiple choice test can result in higher scores than a test that requires recalling an answer because multiple choice tests measure recognition which is superior than recall, and recall tests measure recalling information which is more difficult (Anderson, 2015. p. 172).

2.4.2 Memory and decay

Tulving and Craik (2000) state that memory is "the means by which we retain and draw on our past experiences to use that information in the present". It is important to differentiate between the concept of memory and the concept of retention. According to Schmidt and Lee (2005) memory is "the persistance of the acquired capability for the performance", however retention "refers to the persistance or lack of persistance in the performance" p. 433-443. Memory is a theoretical concept that refers to the continued existence of information, but retention is a behavioural concept and an indicator of the loss or gain of memory (Schmidt & Lee, 2005).

There are four types of memory: *sensory memory, working memory* (WM), s*hort-term memory* (STM), and *long-term memory* (LTM). Baddeley (2004) states that *sensory memory* is the means to hold new content and keep it available for a very brief time, but WM refers to holding and interrelating multiple contents of information at the same time over a limited period of time. STM is associated with current processing and is used by the WM to use information received and to perform an action (Baddeley, 1998). LTM is where human beings save their accumulation of the knowledge of the world (Hunt & H. Ellis, 2003). Baddeley (1998) states

that LTM is the memory in which information is stored in the long term, such as syntactic knowledge and vocabulary knowledge.

LTM is essential for language learning in general and for vocabulary learning specifically. *Schema theory* is a way of organising the information in LTM into units which helps in the use of the language as it involves relating features to a concept or an entity (A. Baddeley, 1998). The association of the acts of *eating* or *writing* with an entity like *table* is an example of *schema theory* and involves using categorised information stored in the LTM about the *table* (Field, 2003). Field (2003) states that a reader or listener builds up what is known as *content schema* when they use their world knowledge about the writer or the speaker. Furthermore, he states that using world knowledge can be used as a framework to aid prediction and to remember the spoken or the written language. Most of the information stored in the LTM is unconsciously stored and little information is consciously stored from intentional learning (A. Baddeley, 1998).

Moreover, processes such as word repetition, breaking down, and analysing words and sentences are processes that occur in STM. The definition of the WM overlaps with STM but is different in its underlying meaning. According to previous research, WM consists of two sensory stores: the *auditory store* and the *visual store*. The *auditory store* consists of the central executive and phonological loop, and the *visual store* consists of the visuo-spatial sketch pad (A. Baddeley, 1998). The central executive is an attentional system that involves the comprehension and meaning of language; the phonological loop involves phonology processing; and the visuo-spatial sketchpad is where spatial information occurs and is processed (Baddeley, 1998). Auditory STM is where spoken content is stored, and it consists of the phonological loop in which speech perception and speech production are linked to the passive phonological store and the articulatory control, respectively (Harley, 2008).

Memorising is an important skill to add to the knowledge of a language and to vocabulary knowledge (Hulstijn, 2001; Nation, 2001; Waring, 2004). The phonological loop consists of the phonological store and a rehearsal mechanism (Baddeley, 1998). Spoken language remains for a very short time in the phonological store (i.e., one to two seconds) and in order to be able to memorise the spoken word it moves from the phonological store to the rehearsal mechanism. It will then be repeated by saying the word in the mind (Baddeley, 1998). Human beings rehearse a spoken word by using a voice in their head, which is known as a *subvocal rehearsal* of words, but written words need to be phonologically encoded and this mechanism is important to recode the written word to the phonological code (Field, 2003).

The difference between WM and STM reveals itself while undertaking cognitive tasks (Dosher, 2005). WM allows individuals to pay attention to information and to be able to manage and alter information according to the demands of the cognitive task (Dosher, 2005).

Individuals can hold limited amount of information in their WM. The type of information (words and language) held in the WM, and the amount of attention needed for this piece of information, plays an important role in limiting the capacity of the WM. Finally, the four types of memory are essential for vocabulary learning and retention and it is important for educators to know how to implement vocabulary techniques in relation to the memory type.

Studies in memory have been conducted to measure speed of recognition, not the accuracy of meaning. Yang (1997) conducted a longitudinal study of 29 undergraduate students over five weeks to measure the participants' recognition, translation, and semantic priming of artificial language with the help of computer, however, this study measures the speed of response over accuracy. According to Schmitt (2010), most memory research measures reaction times (RTs) rather than vocabulary knowledge as words were used only as a convenient stimulus.

2.4.3 Rehearsal techniques

Planning the number of repetition and the distribution of rehearsal is effective and can lead to learning and retaining information in the long-term memory. *Spaced learning* in which learners should rehearse small amounts of information in several distributed sessions over a period of time is more efficient than *massed learning* in which learners rehearse larger amounts of information in one session. Furthermore, learners are better able to remember massed vocabulary items (i.e., a great number of new vocabulary items) that are learnt by *spaced repetition* than words learnt by *successive repetition* (i.e., by repeating words in one learning session). Baddeley (1997) differentiates between two types of repetition distributions. The first is *maintenance rehearsal* which involves repeating existing information but not processing it, such as trying to remember direction information when it is just presented. This type of repetition distribution does not lead to long-term learning and does not ensure recall of information. It can only strengthen the recognition of information. Conversely, the second type, *elaborative rehearsal,* refers to trying to connect the new information with information that already exists. This type of activity can lead to long-term learning.

In L2 acquisition, *spaced repetition* has been investigated with children and with adult undergraduate students learning vocabulary, it has also been investigated with native speakers learning vocabulary in their L1. Kornell (2009) investigated *spaced repetition* in undergraduate students learning words in their L1. He wanted to compare *spaced learning* using 20 word cards with *massed learning* using another 20 word cards. The participants were asked to study 20 word/synonym pair word cards two times in four consecutive sessions, and they were asked to

study 20 flashcards eight times in only one session. The results of the post-test in this experiment show that the students perform better on the words learned via *spaced learning* than the words learned via *massed learning*.

In Fitzpatrick et al.'s (2008) longitudinal study (see section 2.3.2), although the subject could successfully retain 219 out of 400 after four weeks of learning, the file card does not have the advantages of word cards mentioned in section 2.5.1. and did not allow for the distribution of cards using the *Leitner system.* This system is named after a German researcher, Sabastian Leitner and is explained in detail in section 2.5.2.

Planning the retrieval and the space time between intervals is essential to aid memorization from word cards (e.g., Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Dempster, 1988; Kang, 2016; Melton, 1970; Waring, 2004). Repeating a new word a number times and distributing this repetition over the number of intervals as needed, depending on the word difficulty, means retaining this word is more likely to occur at the end of the learning process (Godwin-Jones, 2010; Mondria & Mondria-De Vries, 1994). It is not clear how many times words need to be repeated before they are retained as there is no one golden rule for the ideal repetition rate for any word. As addressed in section 2.4.1., researchers suggest various types of spacing for learning new information, yet none of them justify their suggestions of repetitions.

There are two important memory techniques to consider, the first is *distributed practice,* or what is commonly known as the *spacing effect* (Carpenter et al., 2012; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster & Perkins, 1993; Dempster, 1988). *Spacing effect* is the result of purposefully planning and distributing rehearsing time, and it is one the most replicable and independent psychological phenomenon (Nakata, 2008). It refers to the comparison between *spaced learning* and *massed learning* (A. Baddeley, 1997; Rogers, 2017). The *spacing effect* refers to the information distributed by time, unlike *massed learning* (e.g., Rogers, 2017). *Massed learning* is "periods of work that are substantially longer than the amount of rest between trials″ (Schmidt & Lee, 2005. p 333). *Massed learning* is also defined as the succession of learning in one learning session (Logan, Castel, Haber, & Viehman, 2012; Rohrer & Pashler, 2007). *Massed learning* is effective for short-term performance; however, it is not as effective as *spaced learning* in long-term performance (Bjork & Bjork, 2011; Logan, Castel, Haber, & Viehman, 2012).

The length of the gap between learning and rehearsal is known as *lag* (Rogers, 2017). It is important to distinguish the comparison of *spaced learning* and *massed learning* from the comparison of two spaced learning sessions with different gap lengths between learning and rehearsal, the latter is called *lag effect* (Rogers, 2017). Much research in the Cognitive Psychology literature is conducted to study *lag effect*, in other words the effect of different

length of gaps or spacing between learning and rehearsal (Rogers, 2017).

The second memory strategy is *retrieval practice,* or what is commonly known as the *testing effect*. *Retrieval practice* is "an act of calling information to mind rather than rereading it or hearing it. "The idea is to produce 'an effort from within' to induce better retention." (Roediger & Butler, 2011, p. 20). In the *retrieval practice* strategy, the more the learner can recall information successfully the more the learner is going to be able to remember this information (Goossens et al., 2016; Roediger & Butler, 2011; Rowland, 2014). Recalling involves trying to remember information while it is absent and it is better than just seeing the information and revising it (Kang, 2016).

Many researchers investigate different types of spacing between learning information and rehearsing. Some researchers claim to achieve superior retention of information if the intervals between the review of the information and the rehearsal become bigger and bigger every time the learner successfully recalls the information (Cull, Shaughnessy, & Zechmeister, 1996; Dempster & Perkins, 1993; Landauer & Bjork, 1978; Rea & Modigliani, 1985; Storm, Bjork, & Storm, 2010). Rehearsal with bigger intervals, or what is commonly called *expanding conditions,* encourages learners to review information because over time they make fewer mistakes (A. Baddeley, 2004, p. 92), also having bigger intervals allows for the introduction of new information (Bjork, 1988, p. 399). However, if the learner fails to remember the information the interval should become smaller (Landauer & Bjork, 1978). Examples of two rehearsing techniques that apply *spacing effect* and *retrieval practice,* and also apply smaller spacing of intervals and expanding intervals, are the *Leitner system* and Landauer and Bjork's (1978) spacing system.

In a study conducted by Landauer and Bjork (1978), the researchers introduced face-name pairs to learners using three different rehearsal techniques. The first rehearsal technique was *massed condition* (i.e., no break between intervals), the second was *equal conditions* (i.e., equal spacing intervals), and the third was *expanding conditions* (i.e., increasing spacing intervals). In the *massed condition* rehearsal technique, Landauer and Bjork (1978) presented the face-name pair and repeated it consecutively three times. In the *equal conditions* rehearsal technique, three intervening trials are inserted every time the learner remembers the face-name pair successfully. In the *expanding conditions* rehearsal technique, Landauer and Bjork (1978) asked learners to recall the face-name pair with no intervening trial, when it is recalled the intervening trial increases to one, then to three, then to eight intervening trials every time the learner recalls the face-name pair successfully. The findings show that the *massed condition* technique resulted in less learning than the *expanding condition* and *equal condition* techniques. However, *expanding condition* and *equal conditions* were both very good rehearsal systems, but the *expanding condition* outperformed the *equal condition* by 10% on the final test

(Landauer & Bjork, 1978).

A drawback of this type of rehearsal is that if the learner is going to apply this type of interval without the help of a monitor it is difficult to keep track of the intervals and the intervening trials while learning at the same time (Mondria, 2007; Mondria & Mondria-De Vries, 1994). Also, the learners may get mistaken when planning the target words and the intervening trials (Mondria, 2007; Mondria & Mondria-De Vries, 1994). Therefore, Mondria (2007) and Mondria and Mondria-De Vries (1994) suggest the *Leitner system* to overcome the limits of Landauer and Bjork's (1978) strategy as the learners have total control of the process of moving cards, they can concentrate only on learning words as the moving of the cards is very easy to do. Moreover, this system can be used to successfully apply the *spacing effect* and *retrieval practice* (Mondria, 2007; Mondria & Mondria-De Vries, 1994). Also, the interval condition from the beginning of learning and the time of recall plays an important role in the superiority of the *Leitner system* over the rehearsal technique suggested by Landauer and Bjork (1978). Landauer and Bjork's (1978) rehearsal system starts with smaller intervals and the learning process lasts for a short time, but in the *Leitner system,* the rehearsal starts with longer intervals and the learning process continues for a longer time, this guarantees long-term retention.

Repetition of word learning is not effective if the time between the first encounter and the second encounter is very long because it is considered to be like the first meeting with the word, however, repetition is affective when the time between the first encounter and the second encounter is short as the word is still in the mind of the learner (Nation, 2013, p. 108).

2.5 Techniques for Supporting Vocabulary Learning

2.5.1 Word cards

The use of word cards is often criticised because words that are newly learned through this method are easily forgotten (Waring, 2004). However, Nation (2013) argues that word cards allow for fast learning and long-term retention as fast learning does not necessarily mean fast forgetting. Word cards, in their simplest definition, are “a set of cards where the L2 word is written on one side and its L1 translation or definition on the other” (Nakata, 2008, p. 4) (see Figure 3.1 and Figure 3.2. Researchers have shown several successful uses of word cards in the classroom. For example, Westphal (1977) conducted an informal experiment in which he asked French students, whom he had been teaching for two years, to keep a file of index cards. His experiment shows the learner could learn the information on 540 to 900 cards successfully.

Word cards can be a means of learning and retaining vocabulary independently and intentionally (Nation, 2001; Waring, 2004). Learning unknown information using word cards requires conscious concentration from the learner (Mondria & Mondria-De Vries, 1994). Word card memorization is flexible because learners can arrange cards according to their needs by paying more attention to the most difficult and less attention to the easiest words (Nakata, 2008; Waring, 2004).

Elgort's (2007) doctoral study shows evidence of the advantage of the word cards strategy. She asked advanced ESL learners to learn 48 plausible non-words by studying for four hours a week using word cards. In the post-test, which implemented the LDT, the learners recalled the non-words and achieved automaticity in lexical access (as cited in Cobb & Horst, 2011). More evidence for the benefits associated with using word cards can be found in Nakata's (2008) study. He investigated the use of word cards and compared it with the use of word lists, and he also compared *spaced learning* and *massed learning* among 120 Taiwanese EFL participants. He used a questionnaire and conducted interviews to ask learners for their opinion of the word cards and word lists. The overall results show that word cards are superior and more preferable to the learners than are the word lists. However, a major limitation of this study is the limited number of words used in the word cards experiment (i.e., ten items) and the short amount of time given to learning the words (i.e., no more than ten minutes).

Learning by intentional learning techniques in general, and more specifically by word cards, has many advantages because of the unique characteristics of word cards. To design a word card a learner writes a new L2 word on one side of the card, on the reverse side the learner writes information about the L2 word, such as a definition, synonyms or antonyms, its translation, a sample sentence, and collocations. A learner can also draw, or use pictures, of the L2 word on the reverse side of the word cards. To learn and retrieve a list of vocabulary each new word should be written on a single word card to allow the learner to move these cards according to their desire (Waring, 2004). One important advantage of word cards is that they can be used manually in paper form, or they can be computerized by computer programing (Godwin-Jones, 2010; Mondria & Mondria-De Vries, 1994; Nakata, 2008). Whether their design is paper-based or computer-based, word cards are a very effective and efficient technique for learning vocabulary for a number of reasons.

Word cards strengthen learners autonomy (Elgort, 2011; Nation, 2006; Nation, 2013), their use is flexible (Mondria & Mondria-De Vries, 1994), and the systematic movement of the cards can help learners to commit a large number of vocabulary items to long-term memory (Nation, 2013). Word cards allow learners to benefit from a learner-centred classroom (Elgort, 2011; Nation, 2006; Nation, 2013). Allowing students to work independently by being able to personalize cards according to their needs and desire is a student-centred classroom learning

activity (Elgort, 2011; Nation, 2013). In one hour learners are capable of learning between 30 to 100 bilingual word pairs independently (Nation, 1980). Word pairs can be relevant to the word card design by writing one word pair on one side and the other pair on the reverse side of the card (Nakata, 2008; Waring, 2004).

Mondria and Mondria-De Vries (1994) state that learning unknown words using word cards requires the learner's conscious concentration. Paying attention to learning long lists of vocabulary items can result in fast and effective learning of forms, and possibly the context, of new words (A. Baddeley, 1997). It is also beneficial to ask learners to design their own cards as when they experience success in learning new words this boosts their motivation to learn (Mondria & Mondria-De Vries, 1994). Writing down the words on the cards will help learners retain them productively in the long term. In a study conducted by Thomas and Dieter (1987) concerning the effect on vocabulary retention of writing vocabulary while learning, they find that writing vocabulary can help learners to retain vocabulary productively.

The special design of word cards prevents learners from seeing the information on the reverse side of the card and this challenges the learner to try to remember the word before turning the card to the information side (Mondria & Mondria-De Vries, 1994). Learners need to have a sufficient intervening time to recall the word from memory and this can be achieved when using paper-based word cards (Nakata, 2008). The effort of trying to recall a word, and receiving feedback following a delay after seeing the word information, is called *retrieval practice* and, according to Baddeley (1997), results in not only successful recall but also in long term retained vocabulary.

Not only are paper-based word cards better than word lists and notebooks because they allow *retrieval practice*, a result of the need to flip the card to see the meaning or translation, but for many other reasons they make an effective technique for learning vocabulary (Nakata, 2008; Nation, 2013; Waring, 2004). Because paper-based word cards and word lists share similar characteristics, a number of researchers have investigated the differences between them. They argue that learners would benefit more from word cards than from word lists written in one place (Mondria & Mondria-De Vries, 1994b; Nation, 2013; Waring, 2004). Memorising from a list of words written in one place can lead the learner to practising the same order of words each time, thus the learner memorises and recalls the first and last words of the list more effectively than the words in the middle of the list (Lewandowsky, 2005). Whereas with word cards the learners keep changing the sequence of the words and can classify them according to their needs (Nakata, 2008) Additionally, the length of the list has an influence on the memorization of the word as it is more difficult to memorise and recall long lists than shorter ones (Lewandowsky, 2005). Nakata (2008), warns of what is called the *list effect*, as learners might remember some words while memorizing a list, but they cannot remember them when

using these words in real life.

Word cards are flexible because, as mentioned earlier in this section, learners can focus more on the more demanding words and less on the easier words (Nakata, 2008; Waring, 2004). Learners can also classify difficult word cards together and easy word cards together to permit more revision of the difficult words and less revision of the easy words (Nakata, 2008). Word cards allow for retrieval along with *spaced learning*, which is the basis of an important learning theory in the research of memory (A. Baddeley, 1997). Furthermore, flexibility is evident in the use of word cards in the *Leitner system,* a system that organizes word cards for effective retrieval and repetition. A further explanation about the use of word cards in the *Leitner system* is addressed in the following section.

2.5.2 The *Leitner system*

The *Leitner system* allows for intervals; it has been defined as "a deck of cards with a sophisticated repetition system on the basis of ever bigger intervals, guaranteeing optimal long-term retention" (Mondria, 2007, p. 64). Learning words using word cards and the *Leitner system* is a fun way to learn and a real cognitive activity (Mondria & Mondria-De Vries, 1994). Furthermore, the *Leitner system* implements the notion of bigger intervals in learning which has been encouraged by many researchers (Mondria & Mondria-De Vries, 1994; Schuetze, 2015). *Spaced repetition* and *elaborative rehearsal* were applied in the design of the *Leitner system* when it was created by Sebastian Leitner. The Leitner box is designed with five sections in which learners move word cards in the box in a certain way to facilitate *spaced repetition* (Godwin-Jones, 2010) (see Figure 3.3).

The *Leitner system* allows for planning the repetition of word cards and for reducing the time spent learning by not repeating already learned words. The *Leitner system* (called *die Lernkartei* in German) is a study system developed in the 1970s by German researcher Leitner based on Ebbinghaus' (1913) and Pimsleur's (1967) ideas of spaced learning (as cited in Godwin-Jones, 2010; Schuetze, 2015). Godwin-Jones (2010) and Schuetze (2015) state that the *Leitner system* consists of five to six boxes (or five to six sections in a box) to classify cards according to the need for revision. They elaborate that a learner reviews a word card in the first box and decides to move it one box forward (i.e., to the second box) if the information is recalled. However, if the learner does not recall the information on the card it remains in the first box. When reviewing cards in the second box the learner completes the same process so that once the information on the card is recalled it is moved forward to third box, but if the information on the card is not recalled it is moved back to the first box. Godwin-Jones (2010)

explains moving the cards similarly but he suggests that when the learner does not know the word card, s/he keeps it in the same box (section). The learner follows the process of moving cards forwards and backwards and when all the cards move to fifth or sixth box (section) this indicates the end of the process and it means that the learner has retained the information on the cards placed in the last box (section). When one of the boxes (sections) ars filled with word cards the learner should increase the number of repetition to avoid accumulation of word cards in one box (section). Mondria and Mondria-De Vries (1994, p52) summarizes the card movement as follows, "whatever is "known" goes into the next compartment; anything "not known/no longer known" should go back into compartment 1".

If the *Leitner system* is applied using five or six sections in a box, rather than individal boxes, the learner applies the same procces of learning. Mondria and Mondria-De Vries (1994) elaborate on the sections' design. They suggest specific sizes of the sections in the *Leitner system* where the first section is one centimetre wide, the second is two centimetres wide, the third five centimetres, the fourth eight centimetres, and finally, the fifth section is fourteen centimetres wide. The increasing size of the sections allows for automaticity as the learners focus more on learning from the word cards rather than on when to relearn an item. The learner revises information at bigger intervals automatically without having to plan for the size of the interval. The box is designed this way to allow the new cards to move forward from section one, with capacity for a maximum of 3040 cards, to sections two, three, four, and then five, and then to keep the old learned cards in section five, which is the largest section. When the first section has only three cards remaining, it can be filled with 3040 new cards and the process can continue. As illustrated earlier, the cards should be moved forward to the second section, then to the third, then to the fourth, and finally to the fifth every time the word card information is recalled successfully. Words that are not recalled should be moved back to section one. Mondria and Mondria-De Vries (1994) suggest an addition to the *Leitner system* with the fifth section to be a reference for the learners' knowledge. They advise learners to occasionally select a sample of ten word cards from the fifth section to test their knowledge. The number of cards suggested for the first section (i.e., 3040) in the *Leitner system* is discouraging for the learners as they have to go through all 3040 cards in order to be able to move some of them to section two, therefore, they suggest fewer number of cards, namely fifteen to twenty cards as a warmup (Mondria & Mondria-De Vries, 1994).

Mondria and Mondria-De Vries (1994) state that moving cards forward allows the learners to concentrate and spend more time on the difficult cards as they are the remaining cards in the first sections. They claim the learner does not have to be concerned about placing a card according to its level of difficulty, all the learner needs to know is if the word is known it should be moved forward, but if it is not it should be moved backwards. As a result, the most

difficult words end up being grouped together and the same will happen with the easiest words. Mondria and Mondria-De Vries (1994) recommend using word cards and the *Leitner system* for all learners at different levels studying different subjects with different degrees of difficulty. Learners who might need to learn "commercial cards" published in large numbers (e.g., from hundreds to 6000) would find it beneficial to use the *Leitner system* to organize the learning of this large number of cards.

In a study conducted by Mondria and Mondria-De Vries (1994) in a number of secondary schools in the Netherlands, the *Leitner system* implementation was successful for learning French and German vocabulary. The *Leitner system* has been used in German schools for a long time and teachers are trained in how to use it (Schuetze, 2015). Publishers Klett and Langenscheidt produced 600 to 2000 items marketing systems in French and English for German learners and it was so successful that it was marketed for a long time (Mondria & Mondria-De Vries, 1994). Recently, manual repetition of word cards is less preferred by researchers and learners as computerised word card repetition is more popular. Nowadays desktop programs and mobile applications that implement word card techniques and spaced learning to study vocabulary (e.g. *Anki*) are used more than the manual Leitner box (Godwin-Jones, 2010, 2011).

Schuetze's (2015) criticism of the *Leitner system* is that it allows the learners to judge for themselves and they might not make the correct judgment. He suggests it would be better to follow Oxford's (1990) time of relearning. However, cognitive psychology literature provides evidence of the positive effect of self-testing and retrieval (Roediger & Karpicke, 2006). The *Leitner system* allows learners to retrieve information as they have to decide where to move the cards according to their success or failure to recall information (Mondria & Mondria-De Vries, 1994), also, flipping the card to see the reverse side for information about the word is another type of retrieval (Nakata, 2008).

As mentioned previously, the methods of using paper-based word cards and using *spaced repetition* algorithms can be computerized (Nakata, 2008, 2011; Nation, 2013). Generally, using computer-assisted vocabulary learning has many advantages over other vocabulary learning techniques. Computer-assisted vocabulary learning can motivate and engage learners with different learning styles (Hulstijn, 2001; Nakata, 2008, 2011; Nation, 2013). It allows for material to be presented in an attractive way and it allows for independent learning (N.C. Ellis, 1995; Nation, 2013). The CALL-program, used by Mondria and Mondria-De Vries (1994), is an old program which implements the *Leitner system*. A modern desktop program and mobile application that implements *spaced repetition* to study vocabulary is *Anki*. This program offers learners the opportunity to design, customize, edit word cards, and includes an audio feature to enable learners to listen to the pronunciation of words (Godwin-Jones,

2011). Another popular word card program that can found as a mobile application is the *Quizlet flashcard system*. Its attractive feature is that users can add definitions of their words and see other users definitions of the same word (Godwin-Jones, 2011). Although there are advantages of computerizing the *Leitner system,* the manual use of the *Leitner system* is useful when facilities of the learning environment do not permit the use of computers.

It is important to note that previous projects in *spaced repetition* have been conducted in the laboratory to control extraneous variables and very little research of spaced repetition had been conducted in the classroom (Hulstijn, 2003; Kang, 2016). Therefore, more research into *spaced repetition* is required.

2.5.3 Graded readers

Graded readers are texts restricted in length with a controlled number of vocabulary items and a controlled grammatical structure (Nation & Wang Ming-tzu, 1999). Graded readers are criticized by some writers who claim that they manipulate the language and do not provide language as it is used in real life, therefore they are not suitable for learners (Nation & Wang Ming-tzu, 1999; Claridge, 2005; Eldridge & Neufeld, 2009; Cobb, 2008). Graded readers, despite the small amount of vocabulary retention that might occur after using them, are an important way to learn vocabulary because learners are offered multiple encounters with new words in different contexts (Day et al., 1991; Dupuy & Krashen, 1993; Horst et al., 1998; Mason & Krashen, 1997; Pitts et al., 1989). Incidental learning that might occur from reading the graded readers is seen to be essential for foreign language learners to help develop vocabulary gains (Waring & Takaki, 2003).

The use of graded readers is also known as *Basal Reading*, specially designed for second language learners it consists of simplified sentences and high frequency vocabulary (Nation & Wang Ming-tzu, 1999; Waring, 1997). Nation and Wang Ming-tzu (1999) and Nation and Waring (1997) state that graded reader programs or schemes consist of texts at different levels of vocabulary and grammatical structure. In these schemes learners with low vocabulary size who are at the beginner language proficiency level start with the high frequency vocabulary level and easiest grammar structure and move forward to more difficult readers when their language and reading skills improve. Nation (2006) suggests three graded reader levels: the high-frequency graded readers are created for extensive reading purposes and cover high frequency words (i.e., 1,000 to 4,000), the mid-frequency graded readers are from the 6,000 word-frequency level, and the low frequency graded readers are from the 8,000 word-frequency

level, according to the BNC. In all these graded readers, the target words are from a lower frequency level.

Reading graded readers is one way to achieve incidental vocabulary learning, although it has stronger effect on vocabulary recognition than on vocabulary recall (Pellicer-Sánchez & Schmitt, 2010). In addition to the advantage of incidental vocabulary learning from graded readers, graded readers can improve the reader's fluency and improve reading skills and comprehension (Nation & Wang Ming-tzu, 1999), it can also increase the speed of lexical access (Waring, 1997). Nation and Wang Ming-tzu (1999) state that although using graded readers is time consuming, they are beneficial for learning grammatical structures and are an enjoyable way of learning. They add that graded readers are the only way that beginner L2 learners can enjoy reading with ease L2 texts that is as real as natural reading texts which are very difficult for second language learners who have low proficiency levels.

The terms graded readers and extensive reading are sometimes used interchangeably, although there are several differences between them. Waring (1997) states the differences between graded readers and extensive reading; graded readers are modified, but extensive reading uses natural texts that can be read by both native speakers and by learners. Graded readers are easier for beginner L2 learners to read, while extensive reading is difficult for learners with low language proficiency. Graded readers have levels that the learners are expected to go through after committing to the task of reading and understanding the whole text, but the purpose of extensive reading is to entertain and motivate the reader to learn the language. It is advised that graded readers should be used in the classroom to ensure teachers can make individual observations. However, extensive reading requires only a lesson at first to know how the text should be read, then learners are free to take the books outside the classroom (Waring, 1997).

Vocabulary knowledge is a significant factor for better reading comprehension, and conversely, reading can lead to further vocabulary knowledge (Laufer, 1996). Many studies in applied linguistics investigate the appropriate vocabulary coverage and density, the number of target word occurrences in graded readers. A large body of literature has investigated the level of vocabulary frequency (i.e., low, mid, high) in graded reader schemes. Such studies include: the coverage of high frequency words in graded reader schemes; the time taken to learn words from the new frequency level; the amount of reading the learner has to do before moving to the natural texts; and, the applicable design of the graded readers (e.g., Cobb, 2007; McQuillan, 2016; Nation, 2014; Waring & Takaki, 2003 and many more).

According to Nation and Wang Ming-tzu (1999), designing graded readers requires control of the length of the texts as well as the vocabulary items and the grammatical structure of the texts. Unfortunately, graded readers have, for a long time, been over-simplified and

written carelessly without consideration of meaning and communication with readers. Therefore, it is important to consider word frequency, grammatical structure, text length, topic (it should be engaging); target words; and word repetition as these factors can affect the efficiency of graded readers' design and programs (Nation & Wang Ming-tzu, 1999).

Laufer (1989; 1996) suggests the appropriate vocabulary coverage and density. Not being able to interpret meaning of a text written in the L2 is an obstacle to the use of L1 reading skills. Laufer (1996) claims that the learner needs automaticity to know the meaning of an unknown word, regardless of the context, in order to reach the threshold vocabulary and, as a result, enable the use of L1 reading skills in L2 reading. She also claims that learners often come across words they do not know, or ones they think they know but which they misinterpret, and that these kinds of words hinder reading comprehension. Moreover, it is not efficient to rely on the reader guessing the meaning from context as there might be a lack of clues from that context. Context redundancy does not ascertain the possibility of guessing new words, and sometimes, even when the clues are in the text readers may not use them because they consist of unfamiliar words (Laufer, 1996). Writers should be aware of these problems and limit the density of the unknown words (Laufer, 1996; Nation, 2014; Waring & Takaki, 2003). The reduction of unknown words, or of the potentially misinterpreted words, will help learners to read smoothly and with better ease of lexical access (Laufer, 1996; Nation, 2014; Waring & Takaki, 2003). Potentially misinterpreted words were referred to by Laufer (1989) as *deceptive transparency* in which the leaner sees a word as familiar but misinterprets it, such as "infallible" which a learner may interpret as "something that cannot fall". Therefore, Laufer (1989; 1996) suggests that the text coverage should be 95% familiar words (i.e., from the 3000 *word families* or 5000 lexical items) and that careful consideration be taken of the 5% target words.

Reducing vocabulary density to 95% or 96% coverage of high frequency words might not be enough to learn unknown words. Learners might ignore the new or unfamiliar words to enjoy reading or to avoid interrupting the flow of ideas while reading, and many researchers are against the explicit highlighting of target words as this characteristic is not found in natural texts (Waring & Takaki, 2003). Some graded readers highlight target words only for the beginner levels and Waring and Takaki (2003) suggest drawing the learners' attention to the target words by emboldening or highlighting might be beneficial.

Repetition of target words increases the chance of learning unknown words (Pellicer-Sánchez & Schmitt, 2010; Saragi et al., 1978; Vidal, 2011). There is a positive correlation between repetition and vocabulary learning most of the time, moreover, repetition of target words in reading is a significant factor for vocabulary learning (Vidal, 2011). There are different numbers of repetitions required in order for an unknown word to be learned from reading. Webb (2007a; 2007b) suggests repeating target words at least ten times to be able to

learn them from reading. Vidal (2011) however, suggests two to three repetitions of target words in order for an unknown word to be learned.

In another significant study by Waring and Takaki (2003), learners have to come across target words eight times to possibly achieve 50% new word form recognition after three months of learning. This can be applied for prompted recognition, but for unprompted form-meaning recognition the chance decreases to 10% to 15%. Waring and Takaki (2003) claim that meeting the target words 18 times would not improve this chance, and even with a 96% coverage rate learners may not learn new words incidentally. Moreover, there are many factors which affect word learning other than repetition (Nation, 2014) and the level of vocabulary retention from reading may decline in the long term (Pellicer-Sánchez & Schmitt, 2010).

Other factors can affect learning from graded readers other than the vocabulary factor (Laufer, 1996; Nation & Wang Ming-tzu, 1999). Factors such as: the learners' mastery of reading skills in the first and second language (e.g., guessing words from context, making inferences, predicting and comprehending the meaning of the paragraph); the learners' ability to use their experiences and background knowledge to understand the text; and the learners' tendency to be engaged with the content of the graded readers (Laufer, 1996; Nation & Wang Ming-tzu, 1999). Although these factors are important for understanding a text, the threshold for reading comprehension is mainly lexical (Laufer, 1996).

Waring (1997) claims that fluent readers tend to remember ideas better than remembering each word that delivers these ideas, and that this is different for L2 learners at beginner level. He also claims that beginner L2 learners read each letter individually and then become able to decode the word and keep it in their WM. The problem is when moving to the second line of a text the beginner reader will forget the words kept in their WM (Waring, 1997) and will not be able to connect the ideas presented in the text (Laufer, 1996). This can create frustration while reading and make the reading slow (Cobb, 2007, 2008). This unpleasant feeling will go away when the learners use simplified texts and read more words which they can keep in their WM. The simplified texts allow the learner to improve and move from decoding word by word to comprehending the text's ideas (Waring, 1997).

Waring (1997) states that reading natural texts for beginners can involve decoding each word, using notes, and translating words by using a dictionary to understand the message of the text. He also states that this intensive reading might help beginner learners to comprehend the idea at the end, but it will not train their eyes to move smoothly from one line to the next and will hinder fluent reading. Learners do not need to be fluent or advanced L2 learners as they can read texts modified to their needs using graded readers (Nation & Wang Ming-tzu, 1999). The texts use simplified structures and illustrations and words that are not used commonly by native speakers (Waring, 1997).

Teachers can create their own graded readers according to their students' needs. Graded readers are modified texts that can be stories based on famous novels, biographies, travel books, and non-fiction texts (Nation & Wang Ming-tzu, 1999). Usually graded readers are 15 to 130 pages long and the level of the targeted learners is noted by the publisher in the classification of their graded readers. An additional problem is that publishers use their own frequency lists (Nation & Wang Ming-tzu, 1999). However, teachers and learners should be warned that graded readers should not be the only means of learning vocabulary as learning new vocabulary from graded readers is not guaranteed (Waring & Takaki, 2003).

To learn target words incidentally from context the number of occurrences and the distribution of the clues to target words in the design of graded readers is essential. There is a greater chance of learning a target word the more often it occurs in a text (Horst et al., 1998) and the closer the repetition of the target words (Nation, 2013).

Long term retention has not been extensively investigated. Few studies investigate vocabulary retention after one to two weeks of experiments, the time which is considered to be suitable for longer-term retention or 'long-term' recall of vocabulary (Godwin-Jones, 2010). For example, most of the studies that investigate graded readers gather data after reading only and overlook long-term learning (Waring & Takaki, 2003).

2.6 Word Characteristics Which Affect Word Processing

Word frequency is probably one of the most important factors which affect word processing. Traxler (2011) states that high frequency words tend to be learned more easily and faster than low frequency words. In addition, he states that word frequency has been shown to affect performance in online and offline tasks. In addition to word frequency, there are a number of factors that impact on the recognition of written and spoken words, such as the length of words (in either letters or syllables), familiarity, concreteness, and the number of orthographic neighbours a word has (Altmann, 1995; Harley, 2008; Laufer, 1996; Warren, 2012; Wilson, 1988). Word similarity to the L1 (cognate effects) will also affect processing (Traxler, 2011).

There are similarities between English and Arabic in the characteristics of the writing system in that graphemes stand for phonemes (Saigh & Schmitt, 2012), although the writing systems are different. Furthermore, word-level processing in English differs from that in Arabic which can affect ESL word-level processing skills (Fender, 2003). For example, initial consonants and certain difficult sounds in English may hinder learning. Arab ESL learners may have problems with English orthography as it is inconsistent (Fender, 2003). Orthographic neighbourhood and phonological similarity factors create further problems for Arab ESL as

they transfer their phonological knowledge of the L1 to the L2; this can be reflected in their spelling and can hinder learning the L2 (Saigh & Schmitt, 2012).

Short, low-frequency words seem to be easy to retain because of their shortness, but the lack of exposure to them makes learning them more difficult (Milton, 2009). High neighbourhood density slows down processing because words that are similar in form compete with each other during language processing.

2.7 Summary of Key Issues

The previous topics feed into the central aim of the current study: intentional and incidental learning using *spaced repetition* via paper based word cards and graded readers in Saudi Arabia. Recent research (e.g., Elgort, 2011; Laufer, 2003; Nation, 2011) has shown that there are high expectations of incidental learning. Laufer (2003) suggests that incidental learning is not suitable for learners at intermediate proficiency level. Furthermore, Nation (2011) encourages learners at intermediate proficiency level to learn vocabulary intentionally. Elgort (2011) states that, unlike incidental learning, intentional learning can lead to implicit and explicit vocabulary learning. Furthermore, Hulstijn (2003) states that retention rates of vocabulary learned under intentional learning conditions are much higher than for incidental learning. Unfortunately, little empirical evidence exists to support the above claims (Elgort, 2011; Nation, 2011). Even though Elgort (2011), Hulstijn (2003), Laufer (2003) and Schmitt (2000) emphasise the importance of intentional vocabulary learning over incidental vocabulary learning for L2 learners, most research on intentional vocabulary learning is not conducted in the classroom, but in controlled laboratory conditions (Hulstijn, 2003).

Despite teaching and learning vocabulary being of interest to many researchers, there is a lack of practical research that provides straightforward examples to follow in vocabulary learning and teaching (Folse, 2004; Nation, 2011). Although many researchers recommend the use of word cards over other vocabulary-learning techniques (Elgort, 2011; Mondria & Mondria-De Vries, 1994; Nakata, 2008), there is a shortage of empirical evidence from the classroom environment in support of this recommendation (Elgort, 2011; Nakata, 2008).

Taking into account the flourishing research on *spaced learning* and *spaced repetition* using flashcards (Nation, 2011; Waring, 2004), few studies have demonstrated that *spaced repetition* is beneficial for foreign vocabulary learning in undergraduate students (e.g. Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bloom & Shuell, 1981). There is strikingly little research on the effects of *spaced repetition* in vocabulary acquisition in the long term (Goossens et al., 2016). Furthermore, there is limited empirical research that compares *spaced learning* using

flashcards in language acquisition research (Nakata, 2008).

Previous studies on *spaced learning* using word cards and the *Leitner system* have had only one, or sometimes just a few, participants (e.g., Fitzpatrick et al., 2008; Nakata, 2008). Not many researchers who investigate spacing and repetition manipulate the time of retrieval in the study period, or the time of the post-test and how that time effects the retention of information (Wahlheim, Maddox, & Jacoby, 2014). Finally, the clear majority of the studies on spaced repetition or graded readers were conducted in the context of English speaking countries, not in Arabic speaking countries. Therefore, more empirical endeavours are needed in the context of Saudi Arabia. For example, studies in spaced learning, such Fitzpatrick et al. (2008), Kornell (2009), and (Kang, 2016), and studies in graded readers, such as Cobb (2007), McQuillan (2016), Nation (2014), and Waring and Takaki (2003), were all conducted in English speaking countries.

With the previous limitations in mind, this study aims to fill the gap which needs empirical studies in intentional vocabulary learning, incidental vocabulary learning, and *spaced learning* using different learning modes. The current study compares intentional learning using word cards with incidental vocabulary learning using graded readers. It is expected that the study will inspire teachers and students in the classroom who are interested in effective vocabulary learning techniques. This study also aims to provide evidence of vocabulary learning by 75 participants of 45 vocabulary items over three weeks of learning. The study aims to close the gap in the lack of research into the *Leitner system* and the appropriate use of word cards and card design.

2.8 Research Questions and Predictions

This study aims to investigate the potential effects of the *spaced repetition* technique on the intentional and incidental learning of vocabulary by comparing results from three groups of Arab ESL learners of English at a Saudi university. The comparison explores the potential effects of rehearsal as well as that of planned vocabulary repetition on the memories of Saudi L2 learners by using word cards in conjunction with the *Leitner system* as well as graded readers. The performances of both the 'word card group' and the 'graded reader group' are also compared with a 'control group' (a group receiving no vocabulary guidance). This study aims to investigate the proportion of words retained by the Arab ESL learners after several distinct time points. It also aims to explore the approximate number of repetitions necessary for intentional learning that might occur from word cards and the *Leitner system* and compare it with incidental vocabulary learning resulting from the use of graded readers.

2.8.1 Pilot study research question

The following are three main research questions with their sub questions.

1. **Which mode of learning results in superior retention after five days of learning?**
   1.1. Which of the three groups (the RMCG, the FDCG, or the GRG) learn and retain more target-words as shown by the post-tests?
   1.2. Which of the three groups (the RMCG, the FDCG or the GRG) react fastest to the target words as shown by the LDT?
2. **Which mode of learning results in superior retention during five days of learning?**
   2.1. How many forward and backward moves does the RMCG and the FDCG make with the *Leitner system?*
   2.2. How many comprehension questions does the GRG answer correctly by the end of each graded readers' texts?
3. **Which word card design group benefits the most from their card design?**

Based on arguments from earlier research, as well as considering the aims and objectives of the current research, the following hypotheses were formulated related to the relative effectiveness of two designs of word cards, the graded readers, as well as *spaced learning* and the *Leitner system* for L2 vocabulary learning:

2.8.2 Pilot study hypotheses

The hypotheses are in the same order as the previous research questions.

Null Hypothesis: intentional learning occurring from word cards and incidental learning occurring from the use grader readers will not differ.

Alternative Hypothesis 1: The use of FD word cards will result in superior retention by comparison with the use of RM word cards and the use of graded readers after the learning period.

1. The FDCG and the RMCG will retain more from the fifteen target-word using FD word cards and RM cards than from using the graded readers.
2. The FDCG and the RMCG will respond faster in the LDT than the GRG.

Alternative Hypothesis 2: The word card groups (i.e., the FDCG and the RMCG) will perform better than the GRG during the learning period.

3. The FDCG and the RMCG will use the *Leitner system* effectively and will move word cards five times or more.

4. The GRG will be able to respond to most of the graded readers' CQs as the text length and difficulty is adjusted to their language proficiency.

Alternative Hypothesis 3: The FDCG will benefit the most from the card design.

2.8.3 Main study research questions

1. **Which mode of learning results in superior retention during and after fifteen days of learning?**
   - 1.1. Which of the two groups (the RMCG or the GRG) learn and retain more target-word as shown by the weekly immediate post-tests and the delayed post-tests?
   - 1.2. Which of the two groups (the RMCG and the GRG) react fastest to the target-word as shown by the weekly immediate post-tests and the delayed post-tests?
   - 1.3. How many forward and backward moves does the RMCG make in the *Leitner system*?
   - 1.4. How many comprehension questions does the GRG answer correctly by the end of each graded readers' texts?
   - 1.5. Do Arab ESL learners' performance decrease when large quantities of vocabulary input are maintained over an extended and consistent period?
     - 1.5.1. How does the retention of target words in lists A, B and C differ in the weekly immediate post-tests from the delayed post-test for the RMCG, the GRG, and the CG?
2. **Do word characteristics affect word retention and learning for the RMCG and the GRG?**
   - 2.1. How does word length and derivational morphemes frequency affect word retention for Arab ESL learners?
     - 2.1.1. How do phonological characteristics affect word retention for Arab ESL learners?
3. **To what extent does repetition and spacing affect learning using word cards and graded readers?**
   - 3.1. What impact does the time and number of repetitions of the 45 target words, 1 to 8 times for the RMCG and 10 times for the GRG, have on word learning and retention?
   - 3.2. Do time of learning and spacing affect learning from word cards and graded readers?

Based on earlier research and the aim and objectives of the current study, the following hypotheses relate to the relative effectiveness of the use of word cards and graded readers as well as *spaced learning* and the *Leitner system* for L2 vocabulary learning:

2.8.4 Main study hypotheses

Null hypothesis: Intentional learning from word cards and incidental learning from graded readers will lead to the same levels of learning and retention of the target-word.

1. **Alternative Hypothesis 1: The use of word cards will result in superior retention by comparison with graded readers.**
   - 1.1. The RMCG will retain more from the forty-five target-words using word cards than the GRG will from graded readers.
   - 1.2. The RMCG will respond faster, as measured by the LDT RT, and will give more stable performance than the GRG.
   - 1.3. The RMCG will use the *Leitner system* effectively and will move word cards five times or more.
   - 1.4. The GRG will be able to respond to most of the graded readers' comprehension questions as the text length and difficulty is adjusted to their language proficiency.
   - 1.5. Arab ESL learners' performance will decrease to some extent when large quantities of vocabulary input are maintained over an extended and consistent period.
     - 1.5.1. Arab ESL learners will perform better in list A than lists in B and C, and better in list B than in list C.
2. **Alternative Hypothesis 2: Difficult word characteristics will cause difficulty in target word retention for Arab ESL learners.**
   - 2.1. Longer words, words with low frequency suffixes, and difficult words affected by their similarity in the list will be harder to learn and retain for Arab ESL learners.
   - 2.2. Words with phonetic characteristics that are difficult for Arab ESL learners will be hard to learn and retain.
3. **Alternative Hypothesis 3: Repetition, spacing and time of learning will affect learning in both groups.**
   - 3.1. Target-words need to be repeated at least five times for the RMCG to be learned and retained. Words need to occur five times in each of the graded readers and twice in a learning period before their form and meaning can be learned and retained.
   - 3.2. *Spaced repetition* and longer time of learning for the RMCG and for the GRG will affect target-word learning and retention.

3 Pilot Study

3.1 Overview

The aim of this pilot study was to investigate the effects of the spaced repetition technique on the retention of vocabulary, by comparing the results from three groups of Arab ESL learners at Albaha University. The comparison explored the potential effects of rehearsal pattern, as well as that of planned vocabulary repetition, on the retention of Arab ESL learners by using two word cards designs and graded readers. The word card designs were RM used by the RMCG, and the FD used by the FDCG. The performances were compared with each other and with the performance of the GRG, who used graded readers only.

A further aim of the pilot study was to investigate the proportion of target words retained by Arab ESL learners after five days of learning, and to explore the approximate number of repetitions necessary for intentional learning from the RM and FD word cards and the *Leitner system* and compare this with vocabulary learning as a result of incidental learning from the graded readers.

This chapter consists of three major sections, the methodology, the results and a discussion that evaluates the research questions of the pilot study. In this section, the post-test reliability test is addressed. This is followed by responses to the pilot study research questions. In the responses to the research questions the LDT AS and RT of the RMCG, FDCG, and GRG were analysed and compared. Following this is an analysis of the GFT scores of the RMCG, FDCG and GRG were analysed and compared. Then the FWMs and the BWMs of the RMCG and the FDCG were analysed, followed by an analysis of the GRG performance on the graded readers' CQs.

3.1.1 Research questions and hypothesis

The pilot study aims to identify the most appropriate type of word cards (RM or FD) to investigate the comparability of intentional and incidental modes of learning in the main study. That is, to compare the number of target words the participants might gain from intentional

learning using word cards in conjunction with the *Leitner system* with the number of target words they might gain from incidental learning using graded readers. The results were subsequently used to inform the design of the main study.

In order to conduct the pilot study, it was necessary to identify suitable Arab ESL learners at the early stages of learning English. According to the Foreign Languages Department at Albaha University, the participants are from the intermediate proficiency levels. Participant selection was achieved by running a battery of tests. The participants' vocabulary size was tested by conducting Nation and Beglar's (2007) VST. WM capacity was tested by the digit span test, this test was designed for the current study as a control because the participants' ability to remember words from the word cards and graded readers should be as equal as possible. These pre-tests are illustrated in the following sections.

The participants' performance on a grammar test given by the Foreign Languages Department at Albaha University was compared and analysed to ascertain whether the participants had similar language proficiency levels. The participants' responses to English language contact questionnaire were also analysed before selection of the groups. The participants identified in this way then had to possess similar vocabulary sizes as measured by the VST. Fifty students from the pre-intermediate level took part in the pre-tests. Only the participants with the same vocabulary size in the VST (i.e., from 2000 to 4000) and who have the same grades in the grammar test (i.e., B and B+) were selected. The selected participants were allocated to one of the following groups: the ready-made word card group (RMCG), the free design word card group (FDCG), and the graded readers group (GRG). Furthermore, based on the VST results, the target word frequency for the word cards and graded readers was selected from the 6000-word frequency level. Selecting target-words from the low frequency level for the pilot study ensured that the participants do not know the target word before doing the experiment.

The pilot study also aimed to investigate if five days was enough for the participants to learn 15 target words and to indicate if any changes were needed for the main study. It is suggested by Mondria and Mondria-De Vries (1994) to have fifteen to twenty cards as at the beginning of learning in the *Leitner system* and then build on this number gradually. The pilot study further aimed to provide information about the most appropriate word card design to use in the main study and about the ways in which the learners work with the five sections in the *Leitner system*. Elgort (2011) suggests that using word cards is effective when learners can personalize their cards based on their needs and desires. Personalizing word cards helped learners pay attention to the written form of the L2 words and facilitates memorizing (Mondria & Mondria-De Vries, 1994). In the following sections of the Methodology Chapter is an illustration of the *Leitner system* design.

To investigate claims about the word card designs, the pilot study applied two card designs. The first was FD, where the participants were given the chance to design their cards, with some restrictions, and the second was a RM design, where the participants were given RM word cards designed by the researcher. The pilot study provided guidance in how to design the graded readers material in the main study. Because learners need 95% coverage of familiar words (Laufer, 1988; Laufer, 1996; Liu & Nation, 1985; Waring & Takaki, 2003), advice is to carefully plan familiar words and target word in graded readers for successful incidental vocabulary learning. If the text is too difficult, successful guessing would be hard to achieve, and if the text does not repeat target words often enough the chance of learning new words would be low (Laufer, 1988; Laufer, 1996; Liu & Nation, 1985). The fifteen target words were inserted into five texts in the graded readers, each word occurred five times in two texts to make up 5% in every text of the graded readers (see Appendix 3 (C) from session 1 to session 5). The following sections illustrate the graded readers' design.

The number of repetitions of a target word in a text is ten in the graded readers. The graded readers were modified to match the participant vocabulary size in the VST, which is from the 3000- to 4000-word frequency band. The graded readers word frequency should not exceed the 4000-word frequency band. In more recent studies, Laufer and Ravenhorst-Kalovski (2010) suggest two texts with different coverage related to the learners vocabulary size. Advanced learners with knowledge of 8,000 word families should read graded readers with the coverage of 98%, but learners with knowledge of 4,000-5,000 word families, similar to the participants in the current study, should read graded readers with the coverage of 95%. Comprehension questions (CQs) were given at the end of each of the graded readers. The CQs were designed to avoid drawing the learners' attention to the target-word explicitly, and to check if they are able to notice and learn each target-word from context. The CQs also gave an indication of word retention while learning from the graded readers (see Appendix 3 (C) from session 1 to session 5).

To measure word retention while learning from the word cards using the *Leitner system* the repetition of word cards was investigated. The forward card movements (FWMs) and the backward card movements (BWMs) in the five sections of the *Leitner system* were analysed. After five days of learning from word cards with the *Leitner system* and the graded readers, the participants were tested on their *receptive knowledge* of the form of the target word using the LDT and their *productive knowledge* of the meaning of the target word using a gap-fill test (GFT).

3.1.1.1 Research questions

The following are three main research questions with their sub questions.

2. **Which mode of learning results in superior retention after five days of learning?**
   3.1. Which of the three groups (the RMCG, the FDCG, or the GRG) learn and retain more target-words as shown by the post-tests?
   3.2. Which of the three groups (the RMCG, the FDCG or the GRG) react fastest to the target words as shown by the LDT?
4. **Which mode of learning results in superior retention during five days of learning?**
   4.1. How many forward and backward moves does the RMCG and the FDCG make with the *Leitner system?*
   4.2. How many comprehension questions does the GRG answer correctly by the end of each graded readers' texts?
5. **Which word card design group benefits the most from their card design?**

Based on arguments from earlier research, as well as considering the aims and objectives of the current research, the following hypotheses were formulated related to the relative effectiveness of two designs of word cards, the graded readers, as well as *spaced learning* and the *Leitner system* for L2 vocabulary learning:

3.1.1.2 Hypotheses

The hypotheses are in the same order as the previous research questions.

Null Hypothesis: intentional learning occurring from word cards and incidental learning occurring from the use grader readers will not differ.

Alternative Hypothesis 1: The use of FD word cards will result in superior retention by comparison with the use of RM word cards and the graded readers after the learning period.

5. The FDCG and the RMCG will retain more from the fifteen target-word using FD word cards and RM cards than from using the graded readers.
6. The FDCG and the RMCG will respond faster in the LDT than the GRG.

Alternative Hypothesis 2: The word cards groups (i.e., the FDCG and the RMCG) will perform better than the GRG during the learning period.

7. The FDCG and the RMCG will use the *Leitner system* effectively and will move word cards five times or more.

8. The GRG will be able to respond to most of the graded readers' CQs as the texts' length and difficulty is adjusted to their language proficiency.

Alternative Hypothesis 3: The FDCG will benefit the most from the card design.

Chapter 4 is dedicated to the pilot study results and discussion. The following is the pilot study design.

3.1.2 Design

The independent variables for this pilot study were the three between-participant modes of learning the RM word cards, the FD word cards, the graded readers, and fifteen target words. The dependent variables included the number of target words recognised in the LDT and produced in the GFT.

3.1.3 Pre-tests

For effective vocabulary testing, words selected from frequency lists for inclusion in a test need to be representative of what learners need to know (Schmitt, 2000). Good selection affects the validity, reliability and practicality of the test: validity means that the test should measure what it claims to measure; reliability means the extent to which the test is consistent and stable over a period of time; and practicality means that the test should be of a suitable length and take a reasonable amount of time (Schmitt, 2000).

The participants received Nation and Beglar's (2007) VST, the digit span test, and a questionnaire about English language contact before the experiment. In addition, a grammar test was collected and analysed. The grammar test was given by an English language professor in the Foreign Language Department at Albaha University to the participants prior to the experiment.

3.1.3.1 Nation and Beglar (2007) Vocabulary Size Test

Nation and Beglar's (2007) VST uses 14,000 English frequency levels based on *word families* occurring in the BNC; it measures the frequency levels which the participants are unfamiliar with, in addition to measuring the estimated overall vocabulary size.

The test gives an estimation of the vocabulary size and measures written receptive vocabulary knowledge in English. It is designed for native speakers and L2 learners. The test is appropriate for the current study for a number of reasons. According to Nation (2012), the

test measures receptive vocabulary knowledge of written words, which is similar to learning target words using word cards. It is also an appropriate measure for the learners reading the graded readers because it measures vocabulary knowledge needed for reading.

It is important to note that the VST does not measure reading skills. Furthermore, the test measures concept knowledge of words to a smaller degree and is designed to measure the basic meaning of words. Neither does this test measure vocabulary knowledge required for listening, writing, and speaking (these are not being considered in the current study).

It is important to motivate the participants to perform to the best of their ability to achieve reliable results. Nation (2012) suggests giving one-to-one oral instruction to eliminate any chance of students not performing to the best of their knowledge. The participants were told that the test consists of fourteen sections. Each section has ten multiple choice questions with a key word written in bold, and participants were asked to circle the letter (a, b, c, or d) which they believed represents the closest meaning to the key word in the question. They were encouraged to read all fourteen sections and to respond to as many sections as possible. They were also informed of the time limit for completing the test (i.e., 40 minutes). The participants were encouraged to ask questions when needed, and they began the test only when they were ready. To calculate the receptive vocabulary size, the learners' total score was multiplied by 100. The test is found in Cobb (2015) (see Appendix 2 A).

3.1.3.2 Grammar test

A grammar test designed by an English professor in the Foreign Languages Department at Albaha University was given to the participants prior to the main study. The purpose of the test was to evaluate students' grammar knowledge at the end of their grammar course. The maximum score that can be achieved on this test is 60.

The test covers various subjects in grammar such as, the use of plural and singular nouns, subject and verb agreement, the use of the articles *a*, *an* and *the*, the use of verb tense. The test consists of seven questions each have a number of sentences to either complete or choose the correct answer. The points are divided fairly depending on the length of the question and the effort needed to answer the question. While the validity of the test could not be evaluated in the context of the current study, it was adopted as it had been used previously in the university and staff at the university considered it to be appropriate for the students at this university (see Appendix 2 B).

3.1.3.3 Digit span test

A digit span test is a simple test designed to measure WM capacity, more precisely storage and cognitive control (Engel de Abreu, Conway, & Gathercole, 2010; Conway, Kane, & Engle, 2003; Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004). A high digit span score is commonly associated with high intelligence test scores, in other words WM capacity has a direct relationship with intelligence (Hornung, Brunner, Reuter, & Martin, 2011). Chuderski (2013) claims that there is a high correlation between WM capacity and fluid intelligence (i.e. fluid intelligence is the ability to reason and solve novel problems (Horn, 1982)). Additionally, there is an inverse relationship between the WM capacity and developmental disorders where people with disorders such as dyslexia and specific language impairment have low digit span scores (Baddeley, Gathercole, & Papagno, 1998; Helland & Asbjornsen, 2004).

As learning from cards and from the graded readers requires the use of WM capacity, learners WM capacity was measured for all participants. For each section, the participants listened to a sequence of numbers and were asked to repeat the sequence of numbers as accurately as they could. The list started with a stimulus set of three numbers and the size increased by one digit at every three trial throughout the test.

The participants were required to repeat the numbers immediately after they were presented. The participants continued taking the test until they were unable to repeat the sequence correctly. The last number of English numerals that the participant remembers correctly is used as that participant's digit span test score on the English digit span test. The last number of Arabic digits the participants remembered correctly is used as the score for Arabic digit span test. Every sequence of numbers was repeated three times and called a unit. The number of numerals in each trial is the score of this unit. Therefore, possible scores were 3, 4, 5, 6, 7 and 8 points for unit 1, 2, 3, 4, 5 and 6 respectively (see Appendix 2 (C)).

3.1.3.4 Questionnaire on participant backgrounds

Prior to the vocabulary learning task the participants completed a questionnaire regarding their contact with Arabic and English inside and outside the University classrooms. The questionnaire was administered to gather information about the nature of language contact the participants had experienced prior to participating in the current study (see Appendix 3 (D)).

3.1.4 Results of the pre-tests

3.1.4.1 Vocabulary Size Test (VST)

On average the RMCG, FDCG, and GRG reached the 3700-word frequency level. In Table 3.1 it can be seen that on average the FDCG reached a higher VST word frequency level, followed by the GRG, then the RMCG.

Table 3.1 *Means and standard deviations of VST frequency level for the three groups*

group	**word card groups**		**GRG(n=10)**
	RMCG (n=10)	**FDCG (n=10)**	
mean	3650	3740	3710
SD	158.11	189.73	191.19

The distribution of VST scores for the RMCG did not differ significantly from the normal distribution (z = .224; p = .168), according to the Kolmogorov Smirnov Z test. The FDCG and the GRG VST scores were also found to be normally distributed, where (z = .170; p = .200), and (z = .221; p = .182), respectively. There were no outliers, and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .522).

A one-way ANOVA was conducted to compare the VST scores of the RMCG, the FDCG, and the GRG. The differences between the RMCG, the FDCG, and the GRG's VST scores were not statistically significant, (F (2, 28) =.646, p = .532).

3.1.4.2 Grammar test

On average the RMCG, the FDCG, and the GRG achieved 44.10 out of 60 points on the grammar test (see Table 3.2).

Table 3.2 *Means and standard deviations of grammar test scores for the three groups*

group	**word card groups**		**GRG (n=10)**
	RMCG (n=10)	**FDCG (n=10)**	
mean	43.90	44.20	44.20
SD	3.242	3.29	3.64

The distribution of grammar test scores for the RMCG did not differ significantly from the normal distribution (z = .214; p = .200), as assessed by the Kolmogorov-Smirnov Z test. The FDCG and the GRG grammar scores were also found to be normally distributed, (z = .202; p = .200), and (z = .129; p = .200), respectively. There were no outliers and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .895).

A one-way ANOVA was conducted to compare the VST scores of the RMCG, the FDCG, and the GRG. The differences between the RMCG, the FDCG, and the GRG's VST scores were not statistically significant, (F (2, 28) =.026, p = .974).

3.1.4.3 Digit span test

Table 3.3. shows the average of sequences of digits in the Arabic digit span test by the RMCG, the FDCG, and the GRG.

Table 3.3 *Means and standard deviations of digit span test in Arabic for the three groups*

group	**word card groups**		**GRG (n=10)**
	RMCG (n= 10)	**FDCG(n=10)**	
mean	7.40	6.80	7.10
SD	.516	.632	.738

As for English digit span test, Table 3.4 shows the average of the sequence of English digits by the RMCG, the FDCG and the GRG.

Table 3.4 *Means and standard deviations of digit span test in English for the three groups*

group	**word card groups**		**GRG (n=10)**
	RMCG (n=10)	**FDCG (n=10)**	
mean	6.10	5.70	6.10
SD	.568	.483	.738

The distribution of Arabic digit span test scores for the GRG did not differ significantly from the normal distribution (z = .254; p = .067), as assessed by Kolmogorov-Smirnov Z test. However, the RMCG and the FDCG Arabic digit span test scores were found to be not normally distributed (z = .381; p <0.001) and (z = .324; p = .004) respectively.

Because the word card groups Arabic digit span test scores were not normally distributed, a Kruskal Wallis H test was run to compare the differences in performance between the RMCG, the FDCG, and the GRG. The Arabic digit span test scores were not statistically significantly different between groups, χ^2 (2) = 4.032, p = .133.

The distribution of English digit span test scores for the RMCG differed significantly from the normal distribution (z = .370; p< 0.001), according to Kolmogorov-Smirnov Z test. The FDCG and the GRG digit span test scores were also found to be not normally distributed, where (z = .433; p< 0.001), and (z = .305; p< 0.001) respectively.

Because the distribution of the English digit span test score for the RMCG, the FDCG, and the GRG was not normal, a Kruskal-Wallis H test was run to determine if there were any differences in English digit span test scores between the RMCG, the FDCG, and the GRG. English digit span test scores were not statistically significantly different between the groups, (χ^2 (2) = 3.112, p = .211).

In summary, the RMCG, the FDCG, and the GRG achieved similar results in the VST test, the grammar test, and the digit span test tests. There was no significant difference between the three groups in the pre-tests measures.

3.1.4.4 Questionnaire

All the participants are female, born in Saudi Arabia, in their first year in the school of English and Literature, and are native Arabic speakers. On average, the age of participants is 19.4 years (SD = .724).

All participants speak Arabic at home, received the majority of their pre-college education in Arabic, and had not been to an English-speaking region for the purpose of studying English. However, 10% of the participants lived in a situation where they were exposed to English. Two of the participants had been to Australia for between one month and six months, and one participant had been to France for a few weeks (i.e., less than a month) for the purpose of tourism.

Seventy-three per cent of participants rate their Arabic language ability as native, 20% as very good, and 6% as good. All the participants had studied Arabic in a formal school setting for twelve years or more.

Sixty per cent of the participants rated their English language ability as good, 13% as poor, and 26% as very good. Eighty-three per cent of the participants studied English for six years and 16% of the participants studied English for between six and ten years.

Table 3.5 shows that 16% of the participants studied English in elementary school for one to two years, but the rest of the participants did not study English in elementary school. All participants studied English in junior high school and in senior high school for more than two years each. Only 10% of the participants studied English in university or college (other than English Literature) for one to two years.

Table 3.5 *Participants time spent studying English at schools and university*

study level	**per cent of participants study in a level**	**duration of study time**
elementary school	16%	one to two years
junior high school	100%	more than two years
senior high school	100%	more than two years
university	10%	one to two years

The second section of this questionnaire is about the participants contact and communication using the English language. Table 3.6 shows that 90% of the participants said they never communicated with native or fluent speakers of English in English. However, only 6% said they had been communicating with a native or fluent speaker of English weekly, and one participant said she had been communicating with a native or fluent speaker of English monthly in the year prior to the start of the semester.

Table 3.6 *Participants' experience of communicating with fluent speakers of English*

experience of English	**number of times**	**percentile of participants/number of participants**
communicate with fluent speaker of English in English prior to the semester	never	90%
	daily	zero

	weekly	6%
	monthly	one participant
	few times a year	zero

Table 3.7 shows that 20% of the participants said they never tried to speak English to their instructor outside of class, while another 20% claimed they tried to speak English to their instructor outside of class a few times a year prior to the semester. Thirty per cent of the participants said they tried to speak English to their instructor outside of class on a weekly basis, and another 30% said they tried to speak English to their instructor outside of class on a monthly basis prior to the semester.

Table 3.7 *Participants' experience of communicating with instructors in English*

experience of English	**number of times**	**percentile of participants/number of participants**
speak English to their instructor outside of class prior to the semester	never	20%
	daily	zero
	weekly	30%
	monthly	30%
	few times a year	20%

Table 3.8 shows that 80% of participants never tried to speak English to friends who were native or fluent speakers of English, while 20% tried to speak English to friends who were native or fluent speakers of English a few times a year prior to the semester.

Table 3.8 *Participant's experience of communicating with native or fluent friends*

experience of English	**number of times**	**percentile of participants/number of participants**
speak English to friends who were native or fluent speakers of English prior to the semester	never	80%
	daily	zero
	weekly	zero
	monthly	zero
	few times a year	20%

Table 3.9 shows that 40% of the participants said they spoke English to classmates prior to the semester on a daily basis, 23% said they did so on a weekly basis, 30% on a monthly basis, and only one participant said she tried to speak to a classmate in English a few times a year prior to the semester.

Table 3.9 *Participants experience of speaking English with their classmates*

experience of English	**number of times**	**percentile of participants/number of participants**
speaking English to classmates prior to the semester	never	zero
	daily	40%
	weekly	23%
	monthly	30%
	few times a year	one participant

Table 3.10 shows that 30% of the participants said they spoke English to strangers whom they thought could speak English on a weekly basis. Forty-three per cent of the participants said they did this monthly, and 26% said they did this a few times a year.

Table 3.10 *Participants' experience of speaking English with strangers*

experience of English	**number of times**	**percentile of participants/number of participants**
speak English to strangers whom the participant thinks they could speak English	never	zero
	daily	zero
	weekly	30%
	monthly	43%
	few times a year	26%

As for speaking English to service personnel (e.g., bank clerks, cashiers), 26% said they never had this experience. However,10% said they spoke English to service personnel on a weekly basis, 26% on a monthly basis, and 36% a few times a year (see Table 3.11).

Table 3.11 *Participants' experience of speaking English to service personnel*

experience of English	**number of times**	**percentile of participants/number of participants**
speak English to service personnel (e.g., bank clerks, cashiers).	never	26%
	daily	zero
	weekly	10%
	monthly	26%
	few times a year	36%

The participants then selected the amount of time they estimated they spent, on average, doing activities in English prior to this semester. As for watching English language television, 16% said they never watched English language television, 43% said they watched it on a daily basis, and 40% said they watched it on a weekly basis (see Table 3.12).

Table 3.12 *Participants' exposure to watching English television*

experience of English	**number of times**	**percentile of participants/number of participants**
watching English language television.	never	16%
	daily	43%
	weekly	40%
	monthly	zero
	few times a year	zero

As for watching movies or videos in English, 33% said they watched English movies on a weekly basis, 50% said they them on a monthly basis, and 16% said they watched them a few times a year (see Table 3.13).

Table 3.13 *Participants' exposure to English movies and videos*

experience of English	**number of times**	**percentile of participants/number of participants**
watching English movies and videos.	never	zero
	daily	zero
	weekly	33%
	monthly	50%
	few times a year	16%

As for reading novels in English, one participant said she reads English novels daily, 33% said they read novels on a weekly basis, 40% said they read novels on a monthly basis, and 23% said they read novels only a few times a year (see Table 3.14).

Table 3.14 *Participants' exposure to English novels*

experience of English	**number of times**	**percentile of participants/number of participants**
reading English novels.	never	zero
	daily	one participant
	weekly	33%
	monthly	40%
	few times a year	23%

As for reading English language magazines, 60% said they never read English magazines, 16% said they read English magazines on a monthly basis, and 23% said they read English magazines a few times a year (see Table 3.15).

Table 3.15 *Participants' exposure to English magazine*

experience of English	**number of times**	**percentile of participants/number of participants**
reading English magazine	never	60%
	daily	zero
	weekly	zero
	monthly	16%
	few times a year	23%

As for reading English language newspapers, 86% said they never read English newspapers and 13% per cent said they read English newspapers a few times a year. As for listening to songs in English, 6% said they never listened to English songs, 23% listened to English songs on a daily basis, 46% on a weekly basis, 16% on a monthly basis, and 6% listened to English songs a few times a year (see Table 3.16).

Table 3.16 *Participants' exposure to English songs*

experience of English	**number of times**	**percentile of participants/number of participants**
listen to English songs	never	6%
	daily	23%
	weekly	46%
	monthly	16%
	few times a year	6%

Thirty per cent of the participants listed communicating with native English speakers on social media (i.e., Instagram, Twitter, and Facebook) as activities for which they commonly used English prior to the semester.

3.1.5 Participants

The participants are Arab ESL learners, Saudi female students studying English Literature for their bachelor's degree in the Foreign Languages Department at Albaha University (Saudi Arabia). The students' acceptance to the Foreign Languages Department was based on their grades in English courses in high school. The participants were students in the third and fourth level (see Appendix 1 (B) for Albaha university Bachelor in English Literature study plan). Fifty students from the pre-intermediate took part in the pre-tests. Only 30 of these, who had the same vocabulary size on the VST (i.e., from 2000- to 4000-word frequency level) and the same grades in the grammar test (i.e., B or B+), were selected to take part in the pilot study. Participants who had less than 2000-word frequency or more than 4000-word frequency in the VST were excluded. Also, participants who had grades less than B- or higher than B+ were excluded. Three participants withdrew from the study for personal reasons. Ten participants were randomly allocated to the three learning conditions: the ready-made word card group (RMCG), the free design word card group (FDCG), and the graded readers group (GRG).

3.1.6 Material preparation

3.1.6.1 Word selection

Words were selected from the 6000-word frequency onward according to the BNC levels. The selection of the 6000-word frequency is a result of the participants' VST scores in the main study. On average, the RMCG, the GRG, and the CG achieved a score of 3700 word frequency level words on the VST. Therefore, to ensure that the learners were not familiar with the target words, the target words were selected from the 6000 word frequency levels and above, i.e., beyond the participants' vocabulary size levels. The number of target words was fifteen and they are listed in Table 3.17

Each text in the graded readers was divided into five short texts (i.e., 500 to 700 word length) to ensure the new words did not exceed 5% of the text, as suggested by Laufer (1989) and Liu and Nation (1985) (see Appendix 3 (C) from session 1 to session 5).

The word *proprietorship* from list A (see Table 3.17), was the only word from the 5,000 word level which was one level higher than the 6000 word frequency level. Also, for the word *postulation* its frequency could not be found in BNC list, but the verb *postulate* is from the 7,000 word frequency level.

Table 3.17 *Target-word list and word token frequency band according to BNC levels* for list A

list A	**frequency**
proprietorship	5,000
contentious	6,000
banish	6,000
shrewd	6,000
concur	7,000
pertinent	7,000
trespass	7,000
contrive	7,000
ingenious	7,000
perceptive	8,000
detestable	8,000
elucidate	8,000
vicissitude	13,000
new-fangled	16,000
postulation	….

3.1.6.2 Word cards

The target English word was typed on one side of the RM word cards (see Figure 3.1), and the L1 translation, synonym, and an example, on the other side (see Figure 3.2). The word cards used in this study are 3 inches by 5 inches, the font size is 36 for the target word and for the information except for the example which is in font size 26. More samples of word cards

can be seen in Appendix 3 (A). In addition to the word cards three records were designed (see Appendix 3 (B)). The FDCG received empty cards and were instructed to create a similar design.

Figure 3.1 Example of the word *vicissitude* on the front of a word card

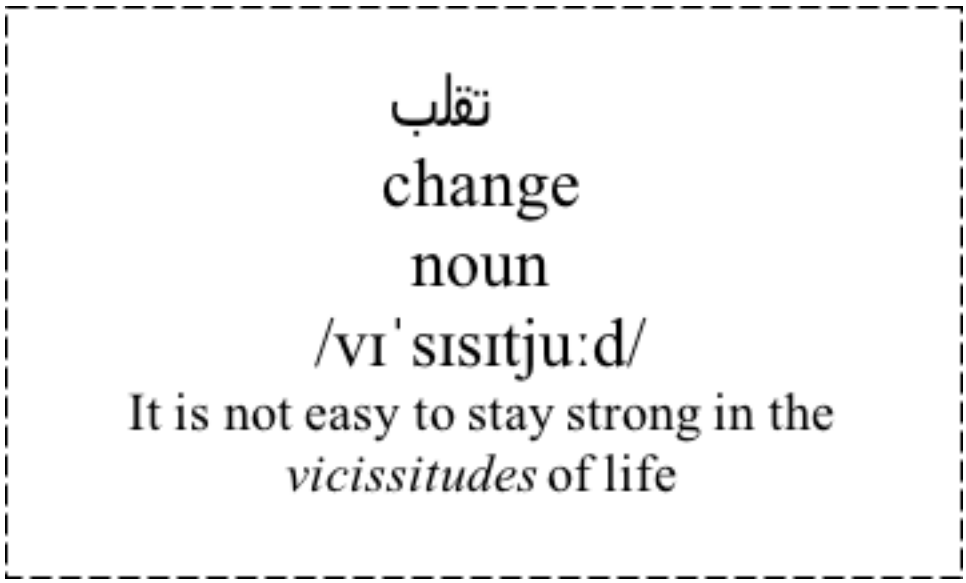


Figure 3.2 Example of *vicissitude* information on the reverse side of a word card

3.1.6.3 The Leitner system

In this study, every participant in the RMCG received a Leitner box. Instead of providing the participants with five boxes, a box with five sections was considered more flexible and economic (see Figure 3.3). Also, information sheets of how to use the *Leitner system* box were

distributed to the word card participants. The system was explained orally and examples of forward and backward movement were illustrated before the start of using the *Leitner system.* Although Mondria and Mondria-De Vries (1994) suggest moving the unknown word card backword to the first section (box) and although Godwin-Jones (2010) suggests to keep the unknown word card in the same section and not moving it, in the current study the participants were asked to move unknown word card backwards one section only. The reason is because moving unknown word card to section one as suggested by Mondria and Mondria-De Vries (1994) might cause frustration for the participants and might slow the learning and not moving it at all as suggested by Godwin-Jones (2010) will prevent the percipients from using the Leitner box fully. It is important to note that the word card group had 20 minutes in each session. Further explanation in section 3.1.7.2.

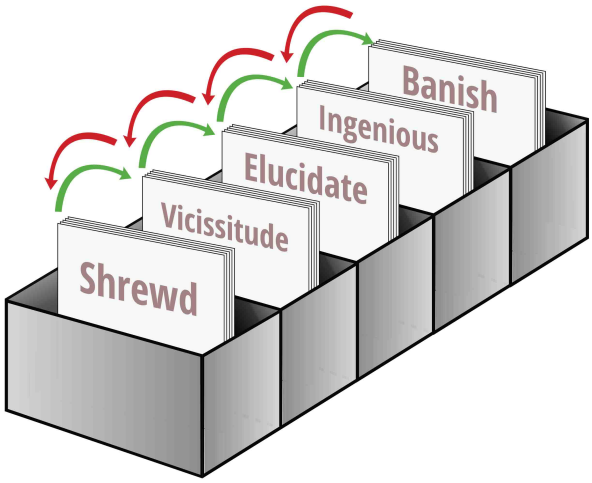


Figure 3.3 Example of the *Leitner system* used for this study

3.1.6.4 Graded readers

The GRG had a reading speed test in which their speed of reading is measured. They were asked to read *A Ghost Story of Christmas,* by *Charles Dickens* from Nation (2014b). This text is a Mid-Frequency Reader adapted to suit readers with a vocabulary of 4000 words, according to the BNC, and participants were asked to draw a line under the sentence they stopped near after 20 minutes reading. The average of the GRG reading speed is from 500 to 700 words in 20 minutes. The target words used in the word card techniques are the same target words used in the graded readers.

The graded readers are suitable to the learners' culture, but are not familiar to them. The text *Free Culture* was chosen from the Graded Readers from Nation's (2014b) website. Appendix 3 (C) (from session 1 to session 5) shows the number of target words in each text and the number repeated in each text. It also shows the length of the texts and the number of CQs.

3.1.6.5 Weekly immediate post-tests and the delayed post-test

The LDT was designed to measure target word recognition, and the GFT to measure target word productive knowledge. The following describes the material preparation for the LDT and GFT.

3.1.6.6 Lexical Decision Task (LDT) material preparation

A Lexical Decision Task (LDT) is a test applied in psychological and psycholinguistic experiments. This test is designed to measure the learner's speed in deciding if a stimulus is a word or a non-word (Elgort, 2011). The LDT involves presenting words and non-words in the centre of a computer screen. The Superlab pro program was used to conduct the LDT. The stimuli were written in upper case and the remainder in lower case. Two instruction screens were designed. The statement on the first screen was written in black, with the word *red* written in red, and the word *blue* written in blue, the instructions read: "Is the word on the screen a real word in English? Press the blue key if "yes" or press the red key if "no". Two statements were written on the second screen: "Place your right-hand finger on the red key and your left-hand finger on the blue key before you start", "Press the space bar when you are ready".

The instruction screen and the stimulus were designed in Microsoft PowerPoint slides and saved as ppt files that were then imported into the Superlab program as pictures. The font for the stimulus was Calibri size 72. Each stimulus was written in black in the centre of a white screen. The participants had to use the J key for "Yes" and the F key for "No" on the keyboard, indicating whether a stimulus was a word or non-word, respectively. The J key was coloured blue and the F key was coloured red.

The stimuli for the LDT consisted of the target word, non-words, and real non-target words. The number of target words was equal to the number of non-words in order to even out the distractor numbers. As for the real non-target words, these were made up of one third of the target word total plus one third of the non-word total. Therefore, the distribution of the LDT

stimuli was 15 target words, 15 non-words, and 10 non-target words (i.e., one third of the target word total is five and one third of the non-word number is five).

The length of stimuli was considered in the LDT design. Stimuli might take longer to read and are more difficult to recognise than shorter words (Field, 2003). The LDT was designed with real words and non-words of different lengths. Table 3.18 shows that the stimuli for words and non-words range from one to six syllables. Although most of the words and non-words had one, two, or three syllables, there was an attempt to add words and non-words with four, five, and six syllables to ensure word length was kept constant between stimuli (see Appendix 4 (B) for the number of syllables in the stimuli).

Table 3.18 *Percentage of stimuli in each syllable count for pilot LDT*

test	**stimuli**	**total number of stimuli**	**number of syllables in a stimulus**					
			1 syllable	**2 syllables**	**3 syllables**	**4 syllables**	**5 syllables**	**6 syllables**
pilot LDT	**Target word**	**15 words**	6.67 %	26.67 %	33.33 %	26.67 %	6.67 %	-
	real word	**5 words**	40.00 %	40.00 %	20.00 %	-	-	-
	non-word	**20 words**	25.00 %	40.00 %	30.00 %	5.00 %	-	-

3.1.6.7 Gap-fill test material preparation

The GFT was designed in the light of the Productive Vocabulary Levels Test (PLT) (Laufer & Nation, 1999). The test was designed to measure productive vocabulary knowledge at different frequency levels to distinguish between learners. The PLT consists of sample sentences designed to elicit knowledge at each frequency level. Each sentence consists of a gap to be filled with a specific word. The sentences provide context to help retrieve the target-word. To target specific word knowledge, the first letters of the missing target-word are provided. To prevent alternative target word choice, the minimum number of letters at the beginning of the gap is added (Laufer & Nation, 1999).

The GFT designed for the pilot study consisted of fifteen gaps. It started with instructions on how to fill the gaps, which consisted of the first two to four letters of the target word, in order to give a clue to the target word to be written in the gaps (see Appendix 4 (C)).

3.1.7 Performance measurement during the tasks

3.1.7.1 The word card performance measurement

Planning for the assessment of the word card movement during the learning sessions was essential. Therefore, the RMCG and the FDCG had to record the movement of word cards. The records were checked at every meeting to ascertain whether the participants understood and applied the instruction accurately (see Appendix 3 (B) for the word records).

3.1.7.2 Section number during the learning session

Every time a participant repeated a word card and decided it was remembered, this word card was moved forward one section. Therefore, the number of FWMs was taken to represent the number of times the participant retained the target word during learning. This measure is subjective because it is based on the participant's own judgment of their knowledge of the target word.

At the end of each learning session, the section number for each word is recorded. This number represented the minimum number of movements of every target-word; minimum because, while learning may had taken 20 minutes, a participant might had moved a word card backward and forward many times before it ended up in its final position. For example, in a learning session, a learner might first move *vicissitude* from section one to section two, then to section three, then to section two again. Therefore, although the learner might have repeated *vicissitude* three times while learning in this specific learning session, the recorded section is two because this is the section in which the card is found at the end of the learning session (see Figure 3.4).

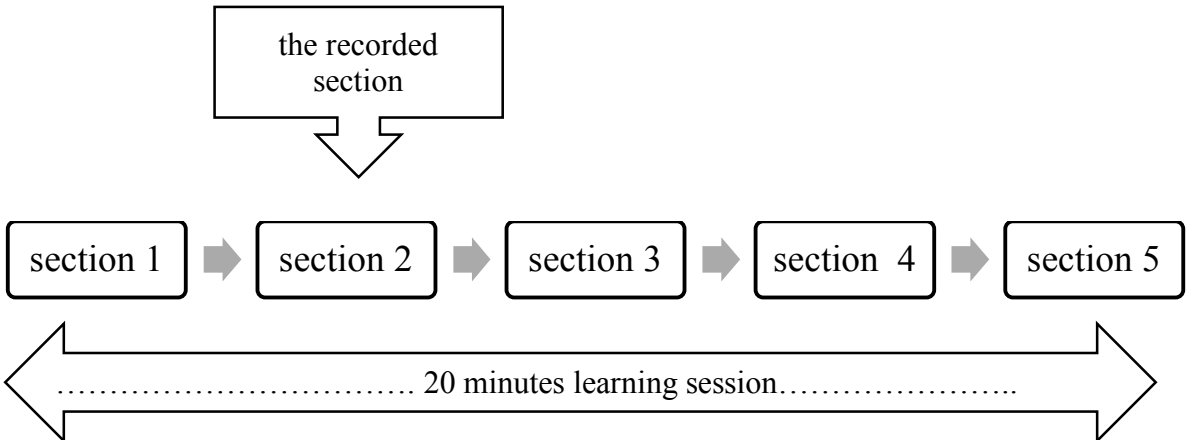


Figure 3.4 Example 1 of recording *vicissitude* word card movement during the first session

If the same learner wanted to rehearse *vicissitude* in the second learning session, she would repeat it by starting from section two because this is the section where the word card was left in the previous session. If the word is remembered, the participant then moves it to section three and then to section four. By the end of the second learning session *vicissitude* would be in section four and the recorded number for repetition would be four because this is where it is found at the end of the second learning session (see Figure 3.5). This process is followed in sessions three to five for every participant.

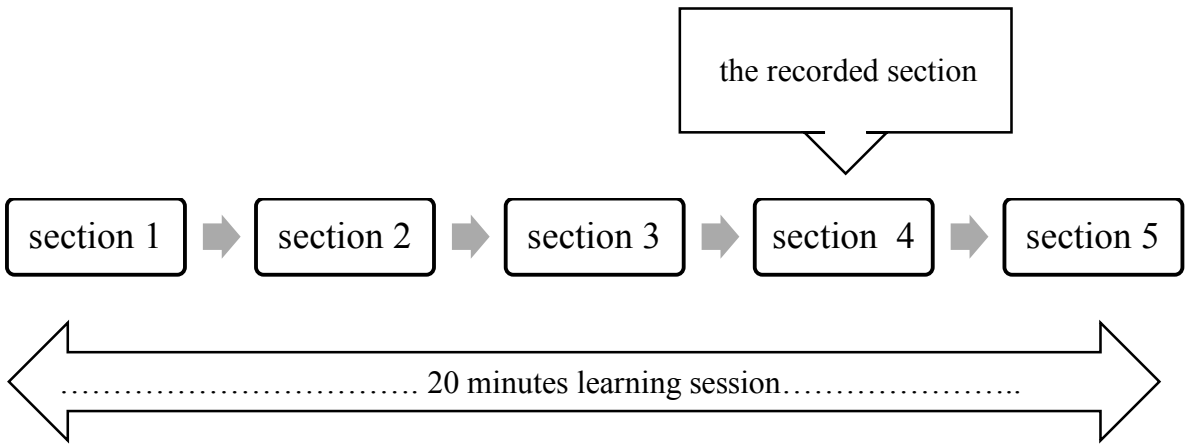


Figure 3.5 Example 2 of recording word card movements for *vicissitude* during the second session

The minimum score is one, because the word cards are first placed in section number one, and the maximum is five. A score of one means that the word card remains in the first section of the *Leitner system,* and the maximum is five because after the word card reaches the fifth section it can no longer be repeated. The process of recording section numbers was repeated in each of the five learning sessions for every word in the 15-word list.

Due to the need for smooth and uninterrupted learning sessions and to the difficulty of reporting all word movements for 15 word cards for ten learners during a 20 minute learning session, the word card movements during the sessions were not recorded. It is more practical to record the number of the last section in which each word card was found at the end of each learning session.

3.1.7.3 The word card movement between learning sessions

At the end of the five sessions, the second step of recording the word card movement consisted of noting the forward movements (FWM) and backward card movements (BWM) for the 15 target words during five days of learning. In this step, two calculations were made: the FWMs calculation and the BWMs calculation. The participants' performances in the *Leitner system* were checked before calculations to ensure that all participants had cards in the *Leitner system* just before the weekly immediate post-tests. In other words, the participants should have at least one card to rehearse in the *Leitner system* before the end of the learning period to ensure that they can be tested immediately after learning takes place.

As the above illustrates, at the end of every five learning sessions each participant left cards in their final places in the *Leitner system.* To measure FWMs every forward movement at the end of every session was calculated as Figure 3.6 shows:

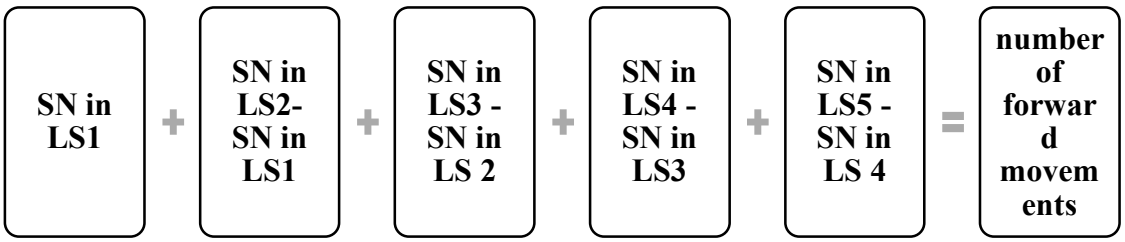


Figure 3.6 Formula for calculating FWMs for every target word in the RMCG
SN = section number LS = Learning session

For example, Sarah have placed *shrewd* in section two by the end of session one, then in section three by the end of session two, in section four by the end of session three, and in section five by the end of session four. In the fifth session, *shrewd* is not rehearsed again because it has reached the fifth section in the fourth session (see Figure 3.7).

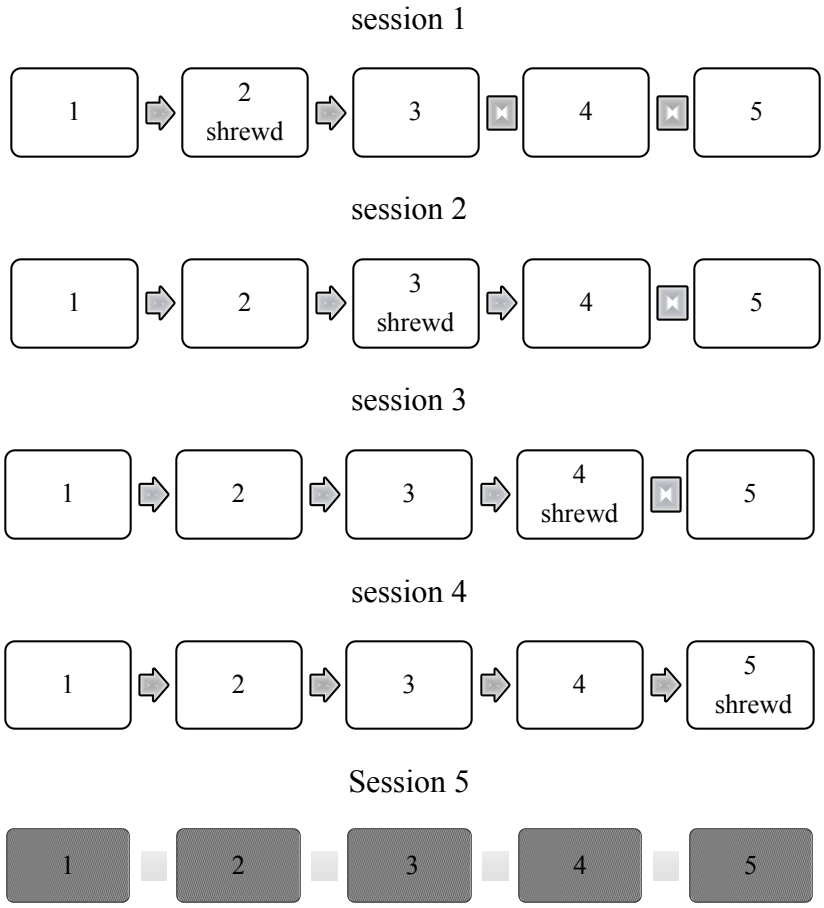
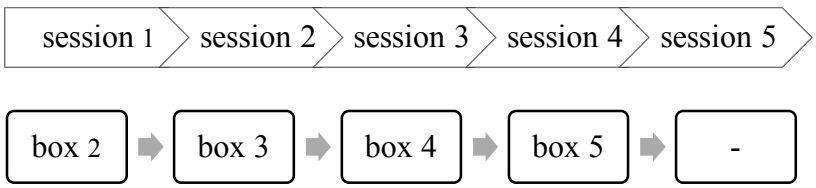


Figure 3.7 Example of FWMs in the five learning sessions.

The forward calculation for Sarah's FWMs of *shrewd* is therefore (2+ (3-2) + (4-3) + (5-4) = 5 forward moves). The box number in session five is not calculated as explained. In this case, Sarah does not move *shrewd* backward anytime between sessions so no backward calculation is performed (see Figure 3.8).



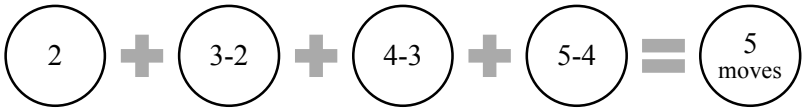


Figure 3.8 Example of FWMs calculation

If BWMs occur during sessions, they are excluded from the FWMs calculation. For example, if Sarah moved *elucidate* across the five sessions to the second section, the first section, the third section, the fourth section, and the fifth section, respectively, the calculation of FWMs would be (2+ (3-1) + (4–3) + (5-4) = 6 FWMs) and the movement from the second section to the first section (i.e., from first session to the second session) is excluded from the FWMs (see Figure 3.9).

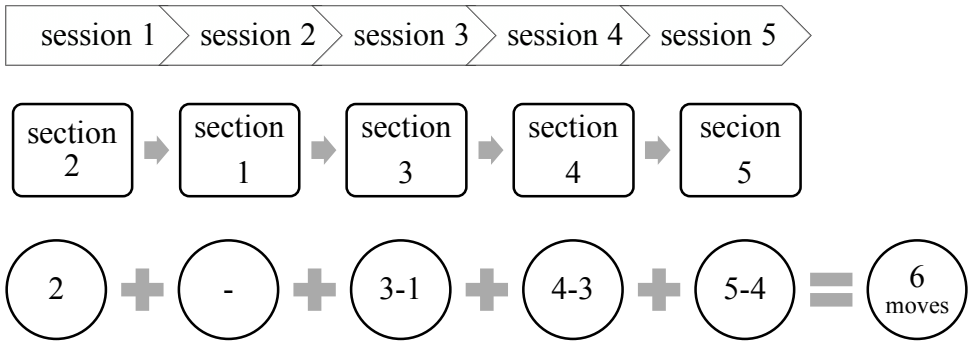


Figure 3.9 Example of FWMs calculation with the exclusion of BWMs

The BWMs need a different calculation. Backward movement shows how many times a participant struggled to remember a target word. To analyse the BWMs, the number of sections where a participant moves cards backward was recorded. For example, if Sarah rehearses *ingenious* in the first session and ends the session with *ingenious* in the third section, then in the second learning session Sarah decides to move it backwards to the second section because she struggles to remember it, this means that she has made one BWM (i.e. 3-2=1). If Sarah then moves *ingenious* to the fourth section in the third session, and then moves it to the second section in the fourth session, and then to the fourth section in the fifth session, this means that in total Sarah has moved *ingenious* backwards three times (i.e. (3-2) + (4-2) = 3) (see Figure 3.10 and Figure 3.11).

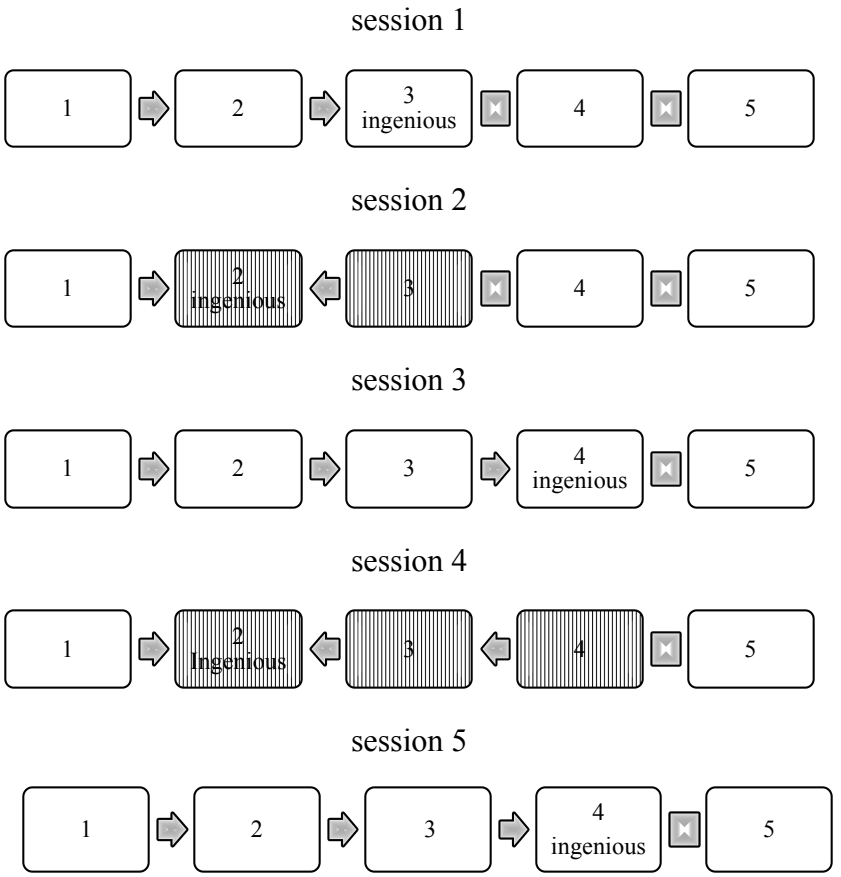


Figure 3.10 Example of BWMS in the five sessions of learning

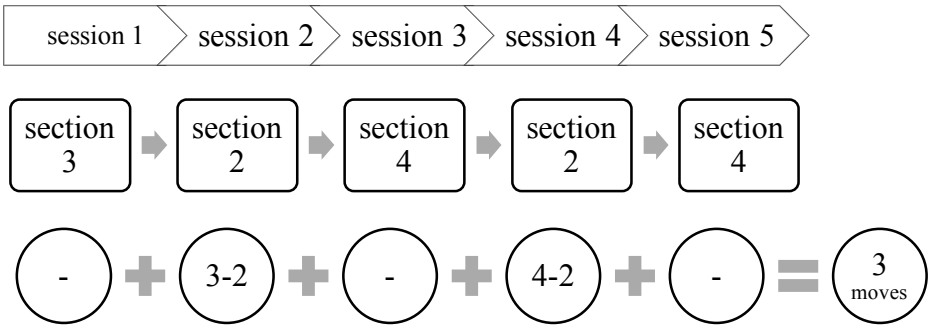


Figure 3.11 Example of BWMs calculation

3.1.7.4 Graded readers' performance measurement

As for the graded readers, the participants' understanding of the target word in the texts was measured according to their responses to the comprehension questions (CQs). A record was designed to keep track of the incidental learning of the target word (see Appendix 3 (E)).

There was a possibility the GRG might not pick up some of the target words, even when reducing vocabulary density to 95% or 96% coverage of high frequency words. To overcome this problem, and to check their understanding of the text, the learners' attention was drawn to the target words by asking CQs at the end of each of the graded readers learning sessions. The CQs measured the GRG ability to infer the meaning of target words and, to some extent, to measure comprehension of the meaning of the context. It also helped draw the learners' attention to target words that they might miss due to speedy reading, or reading for pleasure, and not paying attention to unknown words. There was a CQ for every target word at the end of every text in the graded readers. The CQs do not elicit knowledge of the target word explicitly. The questions ask for information from the text require understanding of the target-word. The CQs were objective measures in which participants had to respond to questions that indicated their understanding of the target word.

The GRG was not allowed to use the dictionary in order to avoid stopping while reading and to move their eyes naturally through the text, as suggested by Waring (1997). Also, the whole purpose of the graded readers was to learn target words incidentally and using the dictionary is contrary this. Every target word had two CQs, as each one occurred in two texts of a graded reader. Therefore, the total number of CQs was 30. The participants were given one point for every correct response. Therefore, the maximum score was 30 (see Appendix 3 (C)) for *Free Culture* texts and CQs.

3.1.8 Post-test measurements

As seen in the Literature Review Chapter, learners' vocabulary knowledge can be measured in terms of *passive* or *receptive* and *active* or *productive* vocabulary knowledge. Anderson and Freebody (1981) suggest other conventions for measuring vocabulary knowledge, which are the *breadth* or *depth* of vocabulary knowledge. *Breadth* of target

vocabulary knowledge in this study is the number of words known from the 45 target words, while *depth* is the degree of understanding of target words (i.e., being able to retrieve it and write it in a sentence), as suggested by Milton (2009), Nation (2001), and Schmitt, (2010). Furthermore, it is important to measure the ease of lexical access to target words. In this study, *receptive* target word knowledge and its *breadth*, along with ease of lexical access were measured using the LDT. *Productive* target word knowledge and its *depth*, in this study, was measured by the GFT. The three groups received two types of tests at the end of each week and after one week of learning. The LDT was designed to measure target word recognition and the GFT to measure target word *productive* knowledge.

When the distribution of one of the groups was not normally distributed, non-parametric tests were used to compare group means. A Kruskal-Wallis H test was run to determine if there were differences in LDT AS, GFT scores or RTs between the RMCG, the FDCG, and the GRG. Effect sizes were computed following the procedure for non-parametric tests described in Field (2013). When the effect size value was between r = 0.20 and r = .50 the effect size was interpreted to be small, and when it was between r = .50 and r = .80 the effect size was interpreted as medium. When the value was r = .80 and above the effect size was considered to be large. Effect size is not computed for correlation significance as according to Field (2013) correlation coefficient is the effect size.

3.1.8.1 Lexical Decision Task (LDT) measurements

Every correct response to a target word was awarded one point in the LDT accuracy scores (AS), and the reaction time (RT) for every correct response only was recorded. An incorrect response to the target word was awarded zero in the LDT AS, and reported as a missing value in the LDT RT. The AS was calculated out of 15 in LDT for each participant in the three groups.

3.1.8.2 Gap-fill test measurements

Incorrect responses to the target word were awarded zero in the GFT and correct responses were awarded one point.

3.1.9 Procedure

The three groups received the pre-tests prior to learning. To ensure that participants' language proficiency levels were similar, the VST and the earlier grammar test, given by the Department of Foreign Languages at Albaha University, were analysed. Also, prior to learning, the participants took a language contact questionnaire and a digit span test. In the learning period the RMCG received RM cards and the Leitner boxes, and the FDCG received empty cards to be designed and the Leitner boxes. The GRG received the graded readers. All three groups were informed of the time limit and the task they were expected to do. They were encouraged to participate as learning by using word cards or graded readers could help them expand their English language knowledge. The GRG were not told that the purpose of the graded readers was to look at their knowledge of vocabulary. The three groups were asked to commit to the study for five weeks, they were motivated to take part because by doing so they receive a certificate in volunteer work from Albaha University. To further motivate participants, they were rewarded with a grade for their hard work in a voluntary study for the University. This grade was agreed on before the main study started as a support for research from Albaha University. After one week of learning the three groups received the LDT and the GFT as a post-test.

Permission was gained from the director of the Foreign Languages Department, as well as from the lecturers of different courses, to allow 45 minutes of lecture time to be spent on the experiment for each group. Fifteen minutes were dedicated to instruction, checking attendance, and discussions, and half an hour for using word cards for the FDCG and the RMCG, and the graded readers for the GRG. They were not informed about the type of test they would take. These sessions took place in a computer laboratory provided by the Foreign Languages Department at Albaha University. Because the LDT was given to the participants immediately after the end of the learning session every week, the last session took place in the computer lab for all groups.

3.1.10 Ethical procedures

Ethical procedures were followed according to the requirements of the Department of English Language and Applied Linguistics at the University of Reading, as well as the Foreign Languages Department at Albaha University in Saudi Arabia (see Appendix 1 (A)).

3.2 Post-tests Reliability Measures

3.2.1 Lexical Decision Task (LDT)

3.2.1.1 LDT accuracy score (AS)

In the pilot study, the second element of measuring word recognition in LDT was the AS for recognition of 15 target words. Every error made was given zero, and 10 points were given for every correct response. The scale had a low level of internal consistency, as determined by a Cronbach's alpha of 0.230. The Cronbach's Alpha was low even with deleted items. Although the scale had a low level of internal consistency, this was an expected result because the participants were trained to learn these target words before the test by using word cards and graded readers.

3.2.1.2 LDT reaction time (RT)

A lexical decision task (LDT) was employed to measure RT of word form recognition. In this pilot study the RTs for fifteen target words were tested. The scale had a high level of internal consistency, as determined by a Cronbach's alpha of 0.879.

3.2.2 Gap-fill test (GFT)

A GFT was given to measure word form and meaning production. The scale had a low level of internal consistency, as determined by a Cronbach's alpha of 0.670. When deleting the items *contentious* and *trespass* the internal consistency became higher with Cronbach's Alpha of 0.702. However, after analysing the learners' performance on the words *contentious* and *trespass* it appeared that the learners repeated them a considerable number of times.

3.3 Evaluation of the Pilot Study Research Questions

To respond to the pilot study research questions, the LDT AS and RT, and the GFT scores of the RMCG, FDCG, and the GRG were compared and analysed. Then the three groups' performance in terms of vocabulary retention while doing the tasks were analysed. In other words, FWMs and BWMs using the *Leitner system* and graded readers' CQs were compared and analysed.

3.3.1 Which mode of learning results in superior retention after five days of learning?

To respond to this research question, an analysis of the three groups' performance in the LDT and the GFT is addressed.

3.3.1.1 Which of the three groups (the RMCG, the FDCG or the GRG) learn and retain more target words as shown by the post-test?

Lexical Decision Task (LDT) (AS)

The three groups took the LDT immediately at the end of five days of learning using the RM word cards, the FD word cards, and the graded readers. The AS target word were recorded. Every correct response to target words was awarded one point, which means that the maximum points participants could obtain was15. The following are the results of the participants' AS. The RMCG had the highest LDT AS, the FDCG came next, and the GRG came last (see Table 4.1).

Table 3.19 *Means and standard deviations of target words LDT accuracy score for the three groups*

group	word card groups		GRG (n=10)
	RMCG (n=10)	FDCG (n=10)	
mean	14.80	14.00	13.40
SD	.42	.94	1.95

The distribution of LDT AS for the RMCG differed significantly from the normal distribution, according to the Kolmogorov-Smirnov Z test, (z = .482; $p < 0.001$). However, the LDT AS of the FDCG and the GRG are found to be normally distributed, where z = .256 and p = .068; and z = .221 and p = .184, respectively.

Because the distribution of one of the groups (i.e., RMCG) is not normally distributed, non-parametric tests were used to compare group means. A Kruskal-Wallis H test was run to

determine if there were differences in LDT AS between the RMCG, the FDCG, and the GRG. The LDT AS do not differ for the three groups. The difference between LDT AS of the three groups is not statistically significantly different, χ^2 (2) = 5.444, p = .066.

Gap-fill Test (GFT)

The GFT was the second test given to the participants after the LDT. In this test, the participants were asked to complete the gaps in 15 sentences. Every gap represents a target word and participants were awarded one point for every correct answer. Therefore, the maximum possible score was 15. The following shows the results of the GFT. The RMCG produced the highest GFT score, the FDCG came next, and the GRG came last (see Table 4.2).

Table 3.20 *Means and standard deviations of gap-fill test score for the three groups*

group	**word card groups**		**GRG (n=10)**
	RMCG (n=10)	**FDCG (n=10)**	
mean	11.60	9.70	6.40
SD	2.50	1.63	2.11

The distribution of GFT scores for the RMCG did not differ significantly from the normal distribution (z = .239; p = .112), according to Kolmogorov-Smirnov Z test. The GFT score of the FDCG and the GRG were also found to be normally distributed, where z = .227 and p = .153; and z = .246 and p = .089, respectively.

A one-way ANOVA was conducted to compare the GFT scores between the RMCG, the FDCG, the GRG. The assumption of homogeneity of variances was met, as assessed by Levene's test of homogeneity of variances (p = .079) The differences between the GFT scores of the RMCG, the FDCG, and the GRG revealed a statistically significant main effect, F (2, 28) =15.46, p < 0.001.

Scheffe post hoc analysis reveals that the mean difference in scores between the RMCG and the GRG was statistically significant (p <0.001). Scheffe post hoc analysis also revealed that the mean difference in scores between the FDCG and the GRG was statistically significant

(p = .007). However, the differences between the RMCG and the FDCG was not statistically significant.

3.3.1.2 Which of the three groups (the RMCG, the FDCG or the GRG) react fastest to the target words as shown by the post-test?

LDT: reaction time (RT)

The GRG was the fastest group to react to LDT target words. The RMCG came next, and the FDCG came last (see Table 4.19). Two RT outliers (6335.80 *ms*) and (6781.87 *ms*) were excluded from LDT RTs of the FDCG and then the mean of the remaining RTs was calculated. The mean RT result from this calculation is (3354.50 *ms*). Two outliers were replaced by the mean RT (3354.50 *ms*) in the FDCG. Subsequently, the mean and standard deviation were calculated after the replacement (see Appendix 5 (A)).

As for the GRG, one RT outlier (6436 *ms*) was excluded in the calculation of the mean of the remaining RTs. After the exclusion of the outlier the mean RT is (2983.51 *ms*). The outlier was then replaced in the LDT RTs of the GRG results by the mean RT (2983.51 *ms*). The mean was calculated out of the LDT RTs of the GRG after the replacement of the outlier.

Table 3.21 *Means and standard deviations of LDT RT for the three groups (ms)*

group	word card groups		**GRG (n=10)**
	RMCG (n=10)	**FDCG (n=10)**	
mean	3104.38	3354.50	2938.51
SD	1264.70	672.45	714.08

The distribution of LDT RTs for the RMCG did not differ significantly from the normal distribution, according to the Kolmogorov-Smirnov Z test (z = 202; p = .200). The LDT RTs of the FDCG and GRG were also found to be normally distributed, where (z = .188; p = .200) and (z = .221; p = .182), respectively. There was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .894).

A one-way ANOVA was conducted to compare the RTs between the RMCG, the FDCG, and the GRG. The differences between the LDT RT of the RMCG, the FDCG, and the GRG were not statistically significant (F (2, 28) =.022, p = .978).

3.3.2 Which mode of learning results in superior retention during the five days of learning?

3.3.2.1 How many forward and backward moves does the RMCG and the FDCG make in the Leitner system?

The method of recording and calculating word card movement is illustrated in the Methodology Chapter. The following are the results of the analysis of the RMCG and the FDCG FWMs (see Appendix 5 (B1)), followed by the results of the analysis of the RMCG and the FDCG BWMs (see Appendix 5 (B2)).

Forward word card movements (FWMs)

Table 3.22 shows that the RMCG had similar FWMs to the FDCG.

Table 3.22 *Means and standard deviations of FWMs for the RMCG and the FDCG*

group	**RMCG**	**FDCG**
mean	5.12	5.12
SD	.164	.097

The distribution of FWMs for RMCG was normal (K-S, z = .184; p = .200). As for FDCG, the distribution of FWMs did not differ significantly from the normal distribution, where (K-S, z = .188; p = .200). There was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .073).

A one-way ANOVA was conducted to compare the FWMs between the RMCG, the FDCG. The differences between the FWMs of the RMCG and the FDCG are not statistically significant (F (1, 19) =.010, p = .922).

Backward word card movements (BWMs)

Table (3. 23) shows that the RMCG average of BWMs was very close to the FDCG average of BWMs.

Table 3.23 *Mean and standard deviation of BWMs for the RMCG and the FDCG*

	RMCG (n=10)	**FDCG (n=10)**
mean	.262	.281
SD	.145	.126

The distribution of BWMs for the RMCG and the FDCG did not differ significantly from the normal distribution, where (z = .178; p = .200) and (z = .235; p = .127), respectively. There was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .677).

A one-way ANOVA was conducted to compare the BWMs between the RMCG and the FDCG. The differences between the BWMs of the RMCG and the FDCG were not statistically significant (F (1, 19) =.097, p = .759).

3.3.2.2 How many comprehension questions does the GRG answer correctly by the end of each graded readers' text?

To explore word retention during the graded reader learning sessions, the mean of scores for comprehension questions were calculated for comparison and analysis. At the end of every graded readers text the participants received a number of comprehension questions for every target word. The total number of CQs was 30 and the participants were awarded one point for every correct answer. Therefore, the maximum score was 30. Every point the participants gained in answering CQs means they had paid attention to this target word. The mean CQ score for the GRG was 24 out of 30 (SD = 1.56).

3.3.2.3 Comparison of the FDCG, RMCG and the GRG performance during learning.

Table 4.6 shows a comparison of the three groups' performance during the task. The number of FWMs for the RMCG and the FDCG and the scores of the GRG in the CQs were reported in percentages. The expected forward moves were 5 times. However, because some learners moved cards more than 5 times (e.g. 6, 7 or 8 times) as a result of moving the cards backwards the mean of FWMs was more than 5 times (i.e. 5.12 times). Therefore, the percentages are higher than 100%.

Table 3.24 *Performance measure across the groups*

	Word card groups		**GRG**
	RMCG	**FDCG**	
mean	102.42%	102. 54%	82.00%
SD	3.292	1.944	5.18

The distribution of FWMs for RMCG was normal, where (K-S, z = .184; p = .200). As for FDCG, the distribution of FWMs did not differ significantly from the normal distribution, where (K-S, z = .188; p = .200). The CQs scores of the GRG were found to be normally distributed, where K-S, z = .180; p = .200, respectively.

A one-way ANOVA was conducted to compare the percentage of word retention while using RM and FD in the *Leitner system* and the graded readers. The assumption of homogeneity of variances was met, as assessed by Levene's test of homogeneity of variances (p = .015) The differences between the percentage of word retention for the RMCG, the FDCG, and the GRG revealed a statistically significant main effect (F (2, 29) = 101.05, p < .0001).

Scheffe post hoc analysis revealed that the mean difference in scores between the FDCG and the GRG was statistically significant (p <.0001). Scheffe post hoc analysis also revealed that the mean difference in scores between the RMCG and the GRG was statistically significant (p < .0001). However, the differences between the RMCG and the FDCG were not significant.

3.3.3 Which word card design group benefits the most from their card design?

To respond to this research question some of the previous results will be highlighted. Pair-wise comparison revealed that the LDT RT between the three groups was not statistically significant. Furthermore, when conducting a Mann-Whitney U test to determine any differences in LDT AS between RMCG and FDCG, the difference between the RMCG and FDCGs was not statistically significant. Additionally, no statistically significant differences were found between the RMCG and the FDCG with respect to the GFT. Moreover, the comparison of the FWMs and BWMs while using the *Leitner system* showed no statistical significance between the two groups.

3.4 Summary and Discussion

In the responses to the research questions, the superiority of the intentional vocabulary learning from FD cards and RM cards over the incidental vocabulary learning from graded readers became clear. Although there was no significant difference between the LDT AS of the FDCG, the RMCG, and the GRG, the word card groups obtained higher scores than the GRG. Moreover, the GFT scores of the RMCG and FDCG were significantly higher than the GFT scores of the GRG.

As for the LDT RT, the GRG was the fastest in responding to stimuli, then RMCG, and FDCG, but being the fastest does not indicate better learning. The FDCG and the RMCG gave higher AS in LDT. According to Elgort (2011), learners who are more aware of the task tend to be slower in their RT, especially when they give high AS. The GRG may have achieved higher AS if they took more time to respond to the target words.

Additionally, the RMCG and FDCG showed better word retention than the GRG during the task. Both the RMCG and the FDCG gave the same performance in the *Leitner system,* indicating the efficiency and practicality of the system. Although the GRG did not show as high word retention as the word card groups, it was considered that learning with the graded readers successful. The CQs played an important role in driving the GRG attention to target word.

Finally, the performance of both the RMCG and the FDCG were very similar during and after the task. However, the RM word card was chosen for the main study, rather than the FD word cards, for two reasons. Firstly, the RMCG obtained slightly higher LDT AS and GFT scores than the FDCG. Secondly, time was limited for the participants in the main study to

design cards for 45 target words and revision of the accuracy and appropriateness of the designs for 25 participants by the researcher would be time consuming. However, RM word cards could be easily controlled by the researcher and would be less time consuming.

4 Methodology

4.1 Overview

The current study was designed to explore the potential effects of spaced intentional learning using word cards and spaced incidental learning using the graded readers on the retention of Arab ESL learners' vocabulary knowledge. Furthermore, the study aimed to explore rehearsal patterns and the approximate number of repetitions necessary for intentional vocabulary retention using word cards and incidental vocabulary retention using graded readers, compared with a control group. A pilot study was required to identify whether ready-made (RM) word cards or free design (FD) word cards were more appropriate for the main study. It also investigated the duration of learning and how repetition in word cards and graded readers can aid vocabulary retention. The pilot study was designed to inform which modes of learning would be appropriate for the main study; it also provides information on the way learners handle the five sections in the *Leitner system*. The battery of tests and the experimental design employed were identical for the pilot and the main study. Therefore, the current chapter sets out the aims and the methodological features common to both and the specific details that are unique to each.

4.2 The Main Study

4.2.1 Design

The independent variables for this main study were: the RM word cards, the graded readers, 45 target words, number of repetitions, and the spacing of learning. The dependent variables include the number of target vocabulary items recognised in the LDT and produced in the GFT. In addition, certain linguistic features of the target words were analysed, including the number of syllables, phonotactic characteristics, and morphological composition. The criteria for identifying these features are reported in the following sections.

In the main study the three groups participating were the RMCG, GRG, and the control group (CG). The CG were not taught vocabulary either explicitly or implicitly. This approach

allows comparison of explicit learning of words using RM word cards with implicit learning from graded readers. As in the pilot study, participants were selected on the basis of a battery of pre-tests.

4.2.2 Pre-tests and results

4.2.2.1 Pre-tests materials

The information about Nation and Beglar's (2007) VST, the digit span test, the questionnaire, and the grammar test is identical to that described in the pilot study, see sections 3.2.3.1; 3.2.3.2; 3.2.3.3; and 3.2.3.4.

4.2.2.2 Analysis of the VST

The distribution of the VST scores for the RMCG, the GRG, and the CG did not differ significantly from the normal distribution (z = .092, p= .188), according to the Kolmogorov Smirnov Z test. There were no outliers, and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .265).

On average the RMCG, the GRG, and the CG achieved 3565.34 points in the VST (see Table 4.1). A one-way ANOVA was conducted to compare the VST scores between the RMCG, the GRG, and the CG. The differences between the three groups' VST scores were not statistically significant, where (F (2, 73) = 1.375, p = .259).

Table 4.1 *Means and standard deviations of VST scores in main study*

group	**RMCG (N=25)**	**GRG (N=25)**	**CG (N=25)**	**overall mean**
mean	3656.00	3580.00	3460.00	3565.33
SD	449.15	442.53	367.42	

4.2.2.3 Analysis of the grammar test

The distribution of the grammar test scores for the three groups differed significantly from the normal distribution, where (z = .178 and p < .0001). The grammar test scores for the GRG did not differ significantly from the normal distribution where (z = .156, p = .118), as assessed by a Kolmogorov-Smirnov Z test. However, the grammar test scores of the RMCG and the CG were not normally distributed, where (z = .176, p = .044) and (z = .214, p = .005), respectively.

There was one outlier in the RMCG (i.e., outlier = 53) and it was replaced by the mean (M = 42.67). The mean 42.67 was the average of scores after removing the outlier. On average, the RMCG, the GRG, and the CG achieved 43.02 out of 60 points on the grammar test (see Table 4.2 for details).

Table 4.2 *Means and standard deviations of grammar test for the main study groups*

group	**RMCG (N=25)**	**GRG (N=25)**	**CG (N=25)**	**overall mean**
mean	42.67	43.24	43.16	43.02
SD	3.48	4.90	4.26	

Given that the RMCG and the CG grammar test scores were not normally distributed, a Kruskal Wallis H test was run to compare the differences in performances between the RMCG, the GRG, and the CG. The grammar test scores were not statistically significantly different between the groups, (χ^2 (2) = .063, p = .969). An assumption was made, on this basis, that the groups were similar in their knowledge of grammar and, therefore, had similar language proficiency.

4.2.2.4 Analysis of the digit span test

The distribution of the Arabic digit span test scores for the RMCG differed significantly from the normal distribution, where (z = .194, p = .016), as assessed by a Kolmogorov-Smirnov Z test. The GRG and the CG Arabic digit span test scores were also found to be not normally distributed, where (z = .256, p < .0001) and (z = .194, p = .016), respectively. On average the three groups could remember a sequence of 6.72 Arabic numbers (see Table 4.3 for details).

Table 4.3 *Means and standard deviations of Arabic Digit Span test scores in the main study*

Group	**RMCG (N=25)**	**GRG (N=25)**	**CG (N=25)**	**overall mean**
Mean	6.68	6.80	6.68	6.72
SD	.98	1.04	.98	

Given that the Arabic digit span test scores were not normally distributed, a Kruskal Wallis H test was conducted to compare the differences in performance between the RMCG, the GRG, and the CG. The Arabic digit span scores were not statistically significantly different between groups, (χ^2 (2) = .357, p = .837).

The distribution of the English digit span test scores for the RMCG differed significantly from the normal distribution, where (z = .324, p = < .0001), as assessed by a Kolmogorov-Smirnov Z test. The GRG and the CG English digit span test scores were also found to be not normally distributed, where (z = .273, p < .0001) and, (z = .318, p < .0001), respectively. On average the three groups remembered a sequence of 5.70 English digits in the English digit span test (see Table 4.4 for details).

Table 4.4 *Means and standard deviations of English Digit Span test scores in the main study*

group	**RMCG (N=25)**	**GRG (N=25)**	**CG (N=25)**	**overall mean**
mean	5.7	5.75	5.67	5.70
SD	.84	.78	.89	

Because the three groups' English digit span test scores were not normally distributed, a Kruskal Wallis H test was conducted to compare the differences in performance between them. The English digit span scores were not statistically significantly different between the groups, (χ^2 (2) = .190, p = .909).

A Mann Whitney U test was conducted to compare the performance of the RMCG, GRG, and CG in the Arabic digit span test scores and in the English digit span test scores. There was a significant difference between the Arabic digit span test scores and the English digit span test scores within the groups (U = 1293.00, z = -5.934, p < .0001). By way of summary, the

RMCG, the GRG, and the CG achieved results that did not demonstrate significant differences in performance on the VST, the grammar test, or the digit span tests.

4.2.2.5 Questionnaire analysis

All the participants are female, born in Saudi Arabia and their first language is informal Arabic. They were in their first year in the school of English and Literature. On average, the participants' age was 19.46 years (SD = .759). All participants speak their L1 at home, they received the majority of their pre-college education in standard Arabic, and none have been to an English-speaking region to study English. However, 19.6% of the participants lived in an environment where they are exposed to English. Table 4.5 shows that some of the participants had travelled to different countries where they could be exposed to English, but they reported that the purpose of such travel was tourism.

Table 4.5 *Percentage of participants who reported their exposure to English outside Saudi Arabia in the main study*

country	**percentages**
did not travel	77.3%
Australia	1.4%
France	4.0%
Malaysia	8.0%
Dubai	4.0%
Bahrain	5.3%
total	100.00%

The participants also reported the length of time they spent outside Saudi Arabia. Twenty per cent of the participants reported they had travelled outside Saudi Arabia for less than a month, and 2.7% had travelled for more than one month but less than six months.

Eighty-four per cent of participants rated their Arabic language ability as native, 13.3% as very good, and 2.7% as good. All the participants had studied Arabic in a formal school setting for twelve years or more. None of the participants rated their English language ability as native-like, but 48% of the participants rated it as very good, 33.3% as good, and 18.7% as poor. As for the duration of studying English, 78.7% of the participants reported studying

English for six years, and 21.3% of the participants had studied English for between six and ten years.

Table 4.6 shows 16% of the participants had studied English in elementary school for one to two years, but the remainder of the participants did not study English in elementary school. All participants had studied English in junior high school and in senior high school for more than two years each. Forty-six percent of the participants studied English in university or college (other than English and Literature) for one to two years.

Table 4.6 *Participants time of studying English in schools and university in the main study*

study level	**percentage**	**duration of study time**
elementary school	9.30%	less than one year
	8.00%	one to two years
	4.00%	more than two years
junior high school	00.00%	less than one year
	00.00%	one to two years
	100.00%	more than two years
senior high school	00.00%	less than one year
	00.00%	one to two years
	100.00%	more than two years
university	53.30%	less than one year
	46.70%	one to two years
	00.00%	more than two years

The second section of this questionnaire relates to the participants contact with, and communication in, the English language. Table 4.7 shows the vast majority of the participants reported that they never communicated with native or fluent speakers of English in English. However, only 10.7% said they had been communicating with native or fluent speakers of English a few times a year, and 8% had been communicating with native or fluent speakers of English monthly in the year prior to the start of the semester.

Table 4.7 *Participants' estimation of number of times they communicated with native or fluent speakers of English*

number of times	**percentage**
never	81.30%
a few times a year	10.70%
monthly	0.00%
weekly	0.00%
daily	8.00%
total	100.00%

Table 4.8 shows that 16% of the participants reported that they never tried to speak English to their instructor outside of class, while a further 25.3% claimed they tried to speak English to their instructor outside of class a few times a year prior to the start of the semester. About 33.3% of the participants said they tried to speak English to their instructor outside of class on a monthly basis, 20% of the participants said they tried to speak English to their instructor outside of class on a weekly basis prior to the semester, and 5.3% reported they spoke to their instructors daily prior to the semester.

Table 4.8 *Percentage of participants for every estimated number of times they engaged in communication with instructors in English outside the classroom prior to the semester*

number of times	**percentage**
never	16.00%
a few times a year	25.30%
monthly	22.30%
weekly	20.00%
daily	5.30%
total	100.00%

Table 4.9 indicates that the vast majority of the participants had never spoken English to classmates prior to the semester. However, 9.3% reported that they spoke to a classmate in

English a few times a year prior to the semester, and 8% reported that they spoke to a classmate in English monthly. Only 1.3% reported they spoke to a classmate in English weekly, and 12% daily, prior to the semester.

Table 4.9 *Percentage of participants for every estimated number of times they engaged in communication with classmates in English outside the classroom prior to the semester*

number of times	**percentage**
never	69.30%
a few times a year	9.30%
monthly	8.00%
weekly	1.30%
daily	12.00%
total	100.00%

As Table 4.10 shows, around three quarters never tried to speak English to friends who are native or fluent speakers of English, while 10.70% tried to a few times a year prior to the semester, and 13.30% said they speak English to friends monthly.

Table 4.10 *Percentage of participants for every estimated number of times they engaged in communication with friends who are fluent or native English speakers in English outside the classroom prior to the semester*

number of times	**percentage**
never	76.00%
a few times a year	10.70%
monthly	13.30%
weekly	00.00%
daily	00.00%
total	100.00%

Table 4.11 shows that 18.7% of the participants never spoke English to strangers who they thought could speak English. However, 37.3% of the participants reported that they did this a few times a year, and 25.3% of the participants reported that they spoke English to a stranger monthly. Sixteen per cent reported that they spoke English to a stranger weekly, and only 2.7% spoke to stranger in English daily prior to the semester.

Table 4.11 *Percentage of participants for every estimated number of times they communicated with strangers whom they thought they could speak English prior to the semester*

number of times	**percentage**
never	18.70%
a few times a year	37.30%
monthly	25.30%
weekly	16.00%
daily	2.70%
total	100.00%

As for speaking English to service industry staff (e.g., bank clerks, cashiers), 12.00% reported that they had never had this experience. However, 48.70% reported they spoke English to service industry staff a few times a year, 22.70% monthly, 13.30% weekly, and 13.30% daily (see Table 4.12).

Table 4.12 *Participants' estimation of number of times they communicated with service industry staff in English prior to the semester*

number of times	**percentage**
never	12.00%
a few times a year	38.70%
monthly	22.70%
weekly	13.30%
daily	13.30%
total	100.00%

The participants then selected the estimated time spent, on average, doing activities in English prior to this semester. The first set of activities involved listening to the English language and familiarising themselves with English culture. As for watching English language television, 4.00% claimed they never watched English language television, 22.70% said they watched it a few times a year, 12% said they watched it monthly, 18.70% weekly, and 42.70% daily.

As for watching movies or videos in English, 4.00% said they watched English movies a few times a year, 2.70% monthly, 92.00% weekly, and only 1.30% daily. As for listening to songs in English, only 4.00% say they never listened to English songs, and 1.30% say they listened to English songs a few times a year. However, 4% listened to English songs monthly, 9.30% weekly, and 81.3% of participants listened to English songs on a daily basis (see Table 4.13).

Table 4.13 *Percentage of participants for every estimated time spent in doing an activity that involves listening to English prior to the semester*

activity	**number of times**	**percentage**
watching English television		
	never	4.00%
	a few times a year	22.70%
	monthly	12.00%
	weekly	18.70%
	daily	42.7%
watching English movies or English videos		
	never	00.00%
	a few times a year	4.00%
	monthly	2.70%
	weekly	92.00%
	daily	1.3%
listening to English songs		
	never	4.00%
	a few times a year	1.30%
	monthly	4.00%
	weekly	9.30%
	daily	81.30%

The second set of English activities on the questionnaire involved participants' reading and their interest in the English culture. As for reading novels in English, 9.3% of the participants said they read English novels a few times a year, 22.7% said they read them monthly, and 68% weekly. As for reading English language magazines, 41.30% said they never read English magazines. However, 5.3% of participants said they read English magazines a few times a year, and 52% said they read English magazines monthly. As for reading English language newspapers, 74.70% of the participants said they never read English newspapers. However, 18.70% of the participants said they read English newspaper a few times a year, 4% monthly, and 2.7% weekly (see Table 4.14).

Table 4.14 *Percentage of participants for every estimated time spent in doing an activity that involves reading in English prior to the semester*

activity	**number of times**	**percentage**
reading English novels		
	never	00.00%
	a few times a year	9.3%
	monthly	22.70%
	weekly	68.00%
	daily	00.00%
reading magazine		
	never	41.30%
	a few times a year	5.30%
	monthly	52.00%
	weekly	1.30%
	daily	00.00%
reading newspaper		
	never	74.70%
	a few times a year	18.70%
	monthly	4.00%
	weekly	2.7%
	daily	00.00%

Sixty-seven per cent of the participants listed communicating with native English speakers on social media (i.e. Instagram, Twitter, Snapchat, and Facebook) as the activities for which they commonly used English prior to this semester.

4.2.3 Participants

The information about the participants is identical to that described in the pilot study methodology, see section 3.2.5., except for the following differences. Seventy-five out of 100 participants were selected. Twenty-Five Arab ESL learners in each group took part in the current study. The RMCG was selected from the Vocabulary Building course group, the GRG was selected from Reading 1 groups, and the CG was from the Grammar 1 course group on the English literature program.

4.2.4 Material preparation

The information about the material preparation is identical to that described in the pilot study methodology, see section 3.2 6.

4.2.4.1 Word selection

The information about word frequency is identical to that described in the pilot study methodology, see section 3.2.6.1, except for the following differences. The total number of target words in the current study was 45. Table 4.15 shows the three target word lists (i.e., list A, list B, and list C), and the token frequency band for each target word. The RMCG were introduced to word list A during the first five days of learning. Word list B was introduced in the second five days in addition to list A. List C was introduced in the third five days of learning in addition to lists A and B.

Table 4.15 *Target word list and word token frequency band according to BNC levels*

list A	**frequency**	**list B**	**frequency**	**list C**	**frequency**
proprietorship	5,000	*unequivocally*	6,000	*folly*	6,000
contentious	6,000	*brittle*	7,000	*feeble*	6,000
banish	6,000	*apprehensive*	8,000	*afflicted*	6,000
shrewd	6,000	*colossal*	8,000	*toil*	7,000
concur	7,000	*idiosyncratic*	8,000	*tedious*	7,000
pertinent	7,000	*misgiving*	8,000	*sluggishly*	7,000
trespass	7,000	*mystification*	8,000	*perseverance*	7,000
contrive	7,000	*obstinate*	9,000	*astound*	7,000
ingenious	7,000	*acrid*	11,000	*lucid*	8,000
perceptive	8,000	*ebullient*	11,000	*grievous*	9,000
detestable	8,000	*stupendous*	11,000	*veritably*	9,000
elucidate	8,000	*gargantuan*	12,000	*petrify*	9,000
vicissitude	13,000	*adroit*	12,000	*cogent*	10,000
new-fangled	16,000	*bonny*	13,000	*expeditious*	12,000
postulation	….	*irksome*	13,000	*rambunctious*	17,000

However, the GRG were introduced to each word list one at a time. The three lists were not mixed due to the difficulty of inserting too many target words into the graded readers (i.e., in five short texts).

In the main study, special attention was paid to consider factors that might influence the recollection of the target words. The duration of articulation relates to the difficulty of memorising target words. Words that take longer duration of articulation are more difficult to be committed to memory than words with shorter duration of articulation (Baddeley, Thomson, & Buchanan, 1975) (see Table 4.16). The syllabifications shown in Table 3.34 are taken from 'How Many Syllables' (n.d.).

Table 4.16 *A classification of target-word and number of syllables*

number of syllables	**1 syllable**	**2 syllables**	**3 syllables**	**4 syllables**
number of target words	**2**	**14**	**16**	**9**
target word	shrewd toil	con.cur con.trive ban.ish tres.pass as.tound ac.rid irk.some co.gent griev.ous bon.ny brit.tle fol.ly feeb.le lu.cid as.tound	con.ten.tious new.fan.gled per.cep.tive in.gen.ious per.ti.nent a.dr.oit stu.pen.dous co.los.sal ob.sti.nate e.bul.lient mis.giv.ing af.flict.ed ram.bunc.tious slug.gish.ly pet.ri.fy te.di.ous	vi.cis.si.tude pos.tu.la.tions de.test.a.ble e.lu.ci.date ap.pre.hen.sive ex.pe.di.tious per.se.ver.ance ver.i.ta.bly gar.gan.tu.an
number of syllables	**5 syllables**		**6 syllables**	
number of target words	**2**		**2**	
target word	pro.pri.e.tor.ship mys.ti.fi.ca.tion		id.i.o.syn.crat.ic un.e.quiv.o.cal.ly	

Also, when selecting a target word, it was important to consider the phonetic features that might cause difficulties for Arab ESL learners (see Literature Review Chapter). The target words were classified according to their phonetic transcription (see Table 4.17). Only target words with the sounds /p/ /v/ and /g/ were considered to a be a potential cause of difficulty for Arab ESL learners as they do not exist in the Arabic language. The sound /g/ might cause

confusion with the sound /k/ and is considered a potential cause of difficulty for Arab ESL learners. The word *cogent* has the sound /dʒ/, an Arab ESL learner might confuse it with the sound /g/. Swan and Smith, (2001) state that the sound /r/ in Arabic is different from the voiced flap /r/ in English, and that Arab speakers tend to over pronounce the post-vocalic /r/, such as in the words *car* and *park*. However, this would not affect the recognition and production of target words because the Arab ESL learners would recognise the letter *r* in the LDT and produce it in the GFT. The consonant clusters and the mirror shape of letters were also considered as a possible reason for difficulties for Arab ESL learners.

Table 4.17 *Target word phonetic transcription and phonological aspect that may cause difficulty for Arab ESL learners*

target word	**phonetic transcription**	**initial two segment consonant cluster (not in Arabic)**	**sounds that do not occur in Arabic**	**mirror shapes letters**
proprietorship	/prəˈpraɪətəʃɪp/	/pr/	/p/	
trespass	/ˈtrespəs/		/p/	
postulation	/ˈpɒstjuleɪeɪʃn/		/p/	
pertinent	/ˈpɜːtɪnənt/		/p/	
stupendous	/stjuːˈpendəs/		/p/	
petrify	/ˈpetrɪfaɪ/		/p/	
perseverance	/ˌpɜːsɪˈvɪərəns/		/p/	
apprehensive	/ˌæprɪˈhensɪv/		/p/	
perceptive	/pəˈseptɪv/		/p/ /v/	
vicissitude	/vɪˈsɪsɪtjuːd/		/v/	
contrive	/kənˈtraɪv/		/v/	
unequivocally	/ˌʌnɪˈkwɪvəkəli/		/v/	
veritably	/ˈverɪtəbli/		/v/	
new-fangled	/ˌnjuːˈfæŋgld/		/g/	
misgiving	/ˌmɪsˈgɪvɪŋ/		/g/	
gargantuan	/gɑːˈgæntʃuən/		/g/	
sluggishly	/ˈslʌgɪʃli/		/g/	
grievous	/ˈgriːvəs/		/g/	
cogent	/ˈkəʊdʒənt/		/dʒ/	

Table 4.17 *Target word phonetic transcription and phonological aspect that may cause difficulty for Arab ESL learners*

target word	phonetic transcription	initial two segment consonant cluster (not in Arabic)	sounds that do not occur in Arabic	mirror shapes letters
detestable	/dɪˈtestəbl/			/d/ and /b/ mirror shape

The phonetic transcription are from 'English Dictionary, Thesaurus, and grammar help | Oxford Dictionaries', n.d.

Another factor worthy of consideration in terms of the target word is the derivational suffix frequency. The number of occurrences in target word lists and the derivational suffixes which occur in each target word is illustrated in Table 3.36. Table 3.36 shows number of type and token frequency (see the Literature Review Chapter for definitions) for each derivational suffix in everyday spoken British English. The values are extracted from the Demographically Sampled component of the BNC (Laws & Ryder, 2014). The words that have suffixes with high frequency are more likely to be learned than target words that have suffixes with low frequency (Nation, 2001). Most suffixes occurred only once in the passages, but to identify a sufficient number of target words some suffixes occurred more than once in the set of derivatives (see Table 4.18).

Table 4.18 *Derivational morphemes frequency and occurrence in target word list*

suffix	PoS of derivative	frequency level	occurrences in target list	Type frequency (target PoS)	Type frequency (other PoS)	Token frequency (target PoS)	Token frequency (other PoS)
-some	(adj)	low frequency	1	14	5 (n)	90	12 (n)
-itude	(n)		1	15		612	
-ship	(n)		1	54		2,366	
-ify	(v)		1	54		1,846	
-ate	(adj form)		1	58	49 (n)	3,645	1,549
-ance	(n)		1	103		6,262	
-ent	(adj)		3	105	44 (n)	9,758	4,817 (n)
-an	(adj)		1	141	98 (n)	3,654	1,119 (n)
-ive	(adj)	mid frequency	2	209	57 (n)	7,193	2,995 (n)
-ous	(adj)		7	261		8,116	
-ation	(n)		1	267		12,570	
-able	(adj)		1	304	16 (n)	6,381	370 (n)
-ate	(verb form)		1	314		10,217	
-ic	(adj)		1	391	81 (n)	7,500	3,296 (n)
-al	(adj)		1	558	181 (n)	37,583	9,853 (n)
-ion	(n)		1	600		52,016	
-ly	(adv)	high frequency	3	1,140	66 (adj)	111,045	8,893 (adj)
-ing	(n)		1	1,425	940 (adj)	32,543	13,665 (adj)
-ed	(adj)		2	1,593	24 (n)	29,633	120 (n)

PoS = part of speech

4.2.4.2 Word cards

The information about word cards is identical to that described in the pilot study methodology, see section 3.2.6.2, except for the following differences. The performance of both the RMCG and the FDCG was very similar during and after the task. However, the RM word card was chosen for the main study, rather than the FD word cards, for two reasons. First, the RMCG obtained slightly higher LDT AS and GFT scores than the FDCG. Second, time was limited for the participants in the main study. Designing cards for 45 target words and revision of the accuracy and appropriateness of the designs for 25 participants by the researcher would have been very time consuming. By contrast, RM word cards could be easily controlled by the researcher and creating these was less time consuming. The word cards were designed and prepared before learning took place in the main study, and they were divided into three sets (A, B, and C) (see Table 3.33). Each set consisted of fifteen word cards.

4.2.4.3 The Leitner System

The information about the *Leitner system* is identical to that described in the pilot study, see section 3.2.6.3.

4.2.4.4 Graded readers

The information about the graded readers is identical to that described in the pilot study methodology, see section 3.2.6.4, except for the following differences. The target words were inserted into three graded readers. Similar to the word cards, participants were introduced to a new list every five days. However, the lists did not overlap. Therefore, list A was inserted into the first graded readers, list B was inserted into the second graded readers, and list C was inserted into the third graded readers. The graded readers' texts were modified for the participant's vocabulary size knowledge (i.e., 3500- to 4000-word frequency), as much as possible.

The texts selected were suitable to the learners' culture and they were not familiar to them. Three texts were chosen from the Graded Readers by Nation (2014b) website. The first text is *Free Culture,* the second is *A Scandal in Bohemia,* and the third is *Metamorphosis*. Appendix 3 (C) shows the number of target word in each text and the number of times it was repeated in each text. It also shows the length of the texts and the number of CQs.

Table 4.19 shows that some target words were repeated consecutively (i.e., the target words were repeated immediately the next day) and some spaced out. The spacing for some words might be one day spacing (i.e, the target words were not repeated the next day immediately but the day after), two days spacing, or three days spacing. This factor was considered in the results of the weekly immediate post-tests and the delayed post-test.

Table 4.19 *The amount of spacing between target words in the five learning sessions for the graded reader*

consecutive repetition	one day spacing	two days spacing	three days spacing
concur	*vicissitude*	*contentious*	*postulation*
new-fangled	*ingenious*	*elucidate*	*banish*
pertinent	*proprietorship*	*adroit*	*stupendous*
detestable	*perceptive*	*gargantuan*	*folly*
misgiving	*shrewd*	*irksome*	*perseverance*
mystification	*trespass*	*toil*	*rambunctious*
obstinate	*contrive*	*tedious*	
colossal	*ebullient*	*afflicted*	
unequivocally	*idiosyncratic*		
veritably	*acrid*		
lucid	*bonny*		
sluggishly	*apprehensive*		
astound	*brittle*		
expeditious	*feeble*		
	grievous		
	cogent		
	petrify		

Furthermore, as Table 4.20 indicates, some target words occurred at different times in the graded readers. Some target words occurred during the first learning session and in the middle of the five days of learning (e.g., *vicissitude* and *contentious),* and other target words occurred at the middle and at the end of the five days of learning (e.g., *elucidate* and *contrive*). Some target words occurred only in the middle of the five days of learning (e.g., *concur* and *new-fangled*), and others occurred at the beginning and at the end of the five days of learning (e.g., *postulation* and *banish*). Some target words occurred near the beginning (e.g., *veritably* and *lucid*) and others occurred near the end (e.g. *expeditious* and *astound*).

Finally, some target words occurred near the beginning and at the end (e.g., *elucidate* and colossal), and others occurred at the beginning and near the end (e.g., *contentious* and *adroit*). Such positioning might had influenced the learning and retention of the target words in the weekly immediate post-tests and the delayed post-test. For example, words that occurred in the third and fifth learning sessions and in the fourth and the fifth sessions might be learned and

retained better than words that occurred in the first and the second sessions and the first and third sessions.

Table 4.20 *The distribution of target words from the beginning, middle, and end of the graded reader five learning sessions*

1st and 2nd	**1st and 3rd**	**1st and 4th**	**1st and 5th**	**2nd and 3rd**
misgiving	*feeble*	*contentious*	*postulation*	*concur*
mystification	*vicissitudes*	*adroit*	*banish*	*new-fangled*
	ingenious	*gargantuan*	*stupendous*	*obstinate*
	proprietorship	*irksome*	*folly*	*colossal*
	ebullient	*toil*	*perseverance*	*veritably*
		tedious	*rambunctious*	*lucid*
				sluggishly
2nd and 4th	**2nd and 5th**	**3rd and 4th**	**3rd and 5th**	**4th and 5th**
idiosyncratic	*elucidate*	*astound*	*contrive*	*unequivocally*
acrid	*afflicted*	*expeditious*	*bonny*	*pertinent*
perceptive			*apprehensive*	*detestable*
shrewd			*brittle*	
trespass			*petrify*	
grievous				
cogent				

4.2.4.5 Weekly immediate post-test and delayed post-test

The information about the weekly immediate post-test and the delayed post-test is identical to that described in the pilot study, see section 3.2.6.5.

4.2.4.6 Lexical Decision Task (LDT) material preparation

The information about the LDT is identical to that described in the pilot study methodolgy, see section 3.2.6.6, except for the following differences. The stimuli were written in capital letters throughout the four LDTs in order to maximize legibility. Non-words like *baldock*, *ralling, moffat, escrotal,* and *oestrogeny* used in the pilot LDT, LDT 1, and LDT 2 were changed to the non-words *flosm, baqule, gambastya, hapgood,* and *jabolk* in LDT 3 and LDT 4. These non-words were changed because they may cause unnecessary delays in processing, which would adversely affected the RT patterns. The reasons the deleted distracters might cause errors are: they are a real word in English (e.g., *abrogative*); they are proper names in English (e.g., *baldock, ralling, moffat)*; or, they are very similar to real words (e.g., *escrotal*

and *oestrogeny*). Motivation for eliminating these non-words comes from a L1 speaker of English who revised the stimuli lists for the LDT. The real words used as distractors in the pilot and the main study LDT were the words *international, help, beautiful, language* and *food.*

The distribution of the LDT 1 stimuli was 15 target words and 15 non-words. As for the real non-target words, the number of stimuli was 10 (i.e., one third of the target word total is five and is added to one third of the non-word number, also five). The number of LDT 2 stimuli was 30 target words and 30 non-words, the real non-target words in LDT 2 was 20 (i.e., one third of the target word total was 10 added to one third of the non-word number, also 10). Finally, the number of the LDT 3 was as follows: 45 target words and 45 non-words. As for the real non-target-word in LDT 3 the number of the stimuli was 30 (i.e., one third of the target word total is 15 added to one third of the non-word number, also 15). The distribution of the stimuli in LDT 3 was similar to the distribution of stimuli in LDT 4. There was an error in this distribution of numbers of real non-target words and non-words in the pilot LDT. The number of real non-target words and non-words was 5 and 20 respectively. This error was solved in the main study (see Appendix 4 (A)).

The length of stimuli was considered in the LDT design. Stimuli might take longer to be read and they were more difficult to be recognised than shorter words (Field, 2003). The LDT was designed with real words and non-words of different lengths. Table 4.21 shows that the stimuli for words and non-words range from one syllable to six syllables. Although most of the words and non-words have one, two, or three syllables, there is an attempt to add words and non-words with four, five, and six syllables to ensure word length was kept constant between stimuli (see Appendix 4 (B) for stimuli syllables).

Table 4.21 *Percentage of stimuli in each syllable count for LDT 1, LDT 2, LDT 3 and LDT 4.*

test	stimuli	total number of stimuli	number of syllables in a stimulus					
			1 syllable	2 syllables	3 syllables	4 syllables	5 syllables	6 syllables
LDT 1	target-word	15 words	6.67 %	26.67 %	33.33 %	26.67 %	6.67 %	-
	real words	10 words	20.00 %	20.00 %	40.00 %	10.00 %	10.00 %	-
	non-words	15 words	6.67 %	80.00 %	13.33%	-	-	-
LDT 2	target-word	30 words	3.33 %	30.00 %	36.6 %	16.67 %	6.67 %	6.67 %
	real words	20 words	25.00 %	50.00 %	15.00 %	5.00 %	5.00 %	-
	non-words	30 words	10.00 %	56.67 %	30.00 %	3.33 %	-	-
LDT 3 and LDT 4	target-word	45 words	4.44 %	31.11 %	35.56 %	20.00 %	4.44 %	4.44 %
	real words	30 words	40.00 %	33.33 %	13.33 %	10.00 %	3.33 %	-
	non-words	45 words	20.00 %	66.67 %	11.11 %	2.22 %	-	-

4.2.4.7 Gap-fill test material preparation

The information about the GFT is identical to that described in the pilot study methodology, see section 3.2.7.7. Four GFTs had been designed for the current study. The first GFT consisted of fifteen sentences. Each sentence had a gap for a target word from list A. The first GFT was used in the pilot study, and also in GFT 1 in the main study because it was designed for the same word list used in the pilot study and in the first learning period of the main study. The second GFT consisted of the same 15 sentences in GFT 1 in addition to 15 sentences testing participants' productive knowledge of word list B. GFT 3 consisted of the 30 sentences in GFT 2 in addition to the 15 sentences designed to test participants' productive knowledge of list C**.** The fourth GFT was similar to GFT 3.

4.2.5 Performance measurement during the tasks

4.2.5.1 The word card performance measurement

The information about the word card performance measurement is identical to that described in the pilot study methodology, see sections 3.2.7.1, 3.2.7.2, and 3.2.7.3, except for the following differences:

4.2.5.2 The word card movement between learning sessions

As illustrated, the participants were introduced to a new list at the beginning of each week of learning. Therefore, the movement of word cards had to be considered during the week they were introduced and during the week that they remained in the *Leitner system* with a new list. Table 4.22 shows how lists A, B, and C were introduced at different times during the 15 days of learning, and words from lists A and B could remain in the *Leitner system* when a new list was introduced.

Table 4.22 *The distribution of lists A, list B, and list C in the three weeks of learning and potential for the continuity of the word card movement in the coming week.*

		first week (list A)					second week (list B)					third week (list C)				
Number of learning session		**Session1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**	**Session 8**	**Session 9**	**Session 10**	**Session 11**	**Session 12**	**Session 13**	**Session 14**	**Session 15**
	Target -word															
Number of box in The *Leitner system*	(list A)	**Box 1**	**Box 1**	**Box 2**	**Box 2**	**Box 1**	**Box 2**	**Box 3**	**Box 4**	**Box 3**	**Box 4**	**Box 4**	**Box 5**	‥	‥	‥
	(list B)	**Not yet introduce**					**Box 1**	**Box 2**	**Box 2**	**Box 2**	**Box 3**	**Box 3**	**Box 4**	**Box 5**	‥	‥
	(list C)	**Not yet introduced**										**Box 1**	**Box 2**	**Box 2**	**Box 3**	**Box 4**

In the process of moving from one week of learning to another, some word cards from previous week(s) were carried over to the following week (e.g., words from list A remained in the *Leitner system* in the second week of learning). In this case, the recording continued and two analyses of FWMs and BWMs were conducted. The first analysis was for the first week of learning only, and the second was for the sessions during which the learner continued to rehearse these word cards during the second week of learning.

For example, if Maha ended the first week with *concur, elucidate,* and *trespass* in the third, third, and fourth sections, respectively, she would continue to record these target words in the second week of learning. Similarly, if Maha had some word cards that did not reach the fifth section from list B during the second week of learning, she would continue recording them in the third week. If Maha could not move some word cards from list A to the fifth section in the first and second weeks, she would continue to record them in the third week.

The mean of FWMs and BWMs were calculated as follows. The number of movements for each word, forwards or backwards, was multiplied by the number of participants who made the movement. The multiplications were summed for each target word then divided by 25 (i.e., the number of RMCG participants). The means were then used for data analysis for FWMs and BWMs. For example, if 10 participants moved the target word *banish* five times forward, 11 participants moved the *banish* 6 times forward, and 4 participants moved *banish* 7 time forward, the number of forward moves for *banish* was (10 x 5) + (11 x 6) + (4 x 7) = 50 + 66 + 28= 144 FWMs. This number of FWMs (144) was divided by 25, giving a mean of 5.76 FWMs. Finally, to show that some of the words were easier for the participants to learn, the target words repeated the most were compared with the least repeated target words in the results analysis.

4.2.5.3 Graded readers' performance measurement

The information about measuring performance while using the graded readers is identical to that described in the pilot study methodology, see section 3.2.7.4, except for the following differences. A record for the 45 target words was designed to keep track of the incidental learning of the target word (see Appendix 3 (E) for records). In every learning week, the participants received a new list of 15 target words (i.e., list A, B, and then C). Every week the GRG received two CQs for each target word. The total number of CQs, and therefore the

possible score, was 90 over the three weeks of learning (see Appendix 3 (C) for graded readers' texts and CQs).

4.2.6 Post-test measurements

The information about the post-test measurement is identical to that described in the pilot study methodology, see section 3.2.8.

4.2.6.1 Lexical Decision Task (LDT) measurements

The information about the LDT measurement is identical to that described in the pilot study methodology, see section 3.1.8.1, except for the following differences. The first LDT measured the participants' recognition of words from list A. The second LDT measured the participants' recognition of lists A and B. The third and fourth LDT measured the participants' recognition of lists A, B, and C.

The outliers for the RTs of correct responses across participants were excluded from the RMCG, the GRG, and the CG before calculating the means and standard deviations for the groups. The mean of the remaining RTs per target word responses was then used to replace the outliers. In Appendix 5 (B) further details can be found regarding the outliers and the means replacing the outliers for the RT for all the groups' results for each target word.

4.2.6.2 Gap-fill test (GFT) measurements

The information about the GFT measurement is identical to that described in the pilot study methodology, see section 3.1.8.2., except for the following differences. The first GFT measured the participants' productive knowledge of words from list A. The second GFT measured the participants' productive knowledge of lists A and B. The third and fourth GFT measured the participants' productive knowledge of lists A, B, and C.

4.2.7 Procedure

The information about the procedure is identical to that described in the pilot study methodology, see section 3.2.9, except for the following differences. In the current study the pre-tests were followed by three weeks of learning: in the first week, the participants were

introduced to list A; in the second week, the participants were introduced to list B; and in the third week they were introduced to list C. In the main study the participants had three weekly immediate post-tests, in each test was the LDT and the GFT. The participants took a delayed post-test after two weeks from learning.

The weekly immediate post-tests assessed participants' vocabulary gains immediately after the learning experience. To illustrate, the participants received the first post-test on list A after the first five days of learning, the second post-test on lists A and B after the second five days of learning, and the third post-test on lists A, B, and C after the third five days of learning. After two weeks and without using any of the learning modes, the RMCG, the GRG, and the CG received the post-test on lists A, B, and C (see Figure 4.1).

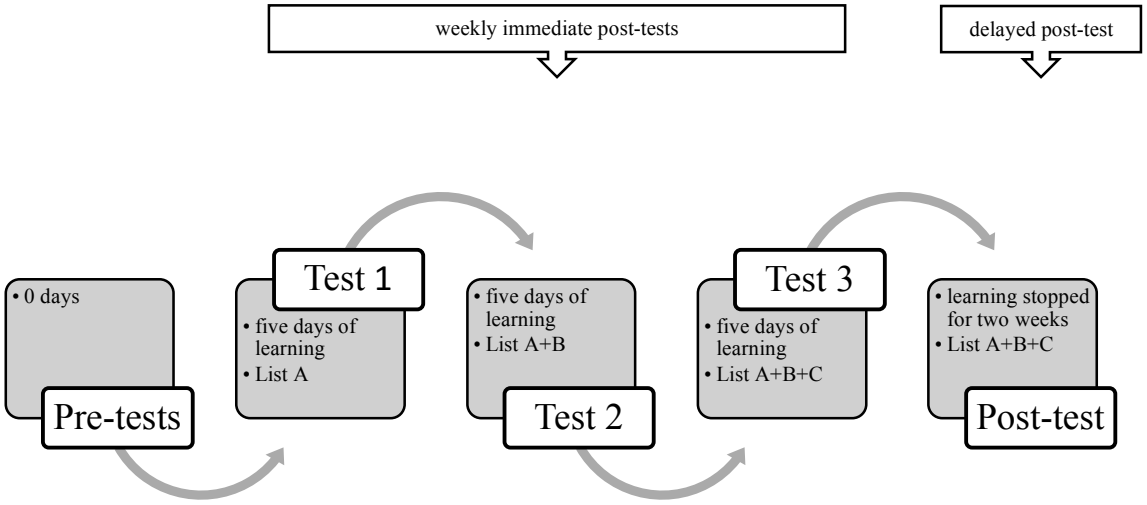


Figure 4.1 The sequence and time of pre-test, weekly immediate post-tests and delayed post-tests.

The first weekly immediate post-test measures participants' knowledge of the first 15 target words. The second weekly immediate post-test measures the participants' knowledge of the second 15 target words plus the first 15 target words (i.e., 30 target words). The third weekly immediate post-test measures the participants' knowledge of the third 15 target words plus the first 15 and the second 15 target words (i.e., 45 target words) (see Table 4.23).

Table 4.23 *Number and distribution of LDTs and GFT over three weeks of learning and at the end of the learning period had ended*

test	weekly immediate post-tests			delayed post-test
time	after 1 week	after 2 weeks	after 3 weeks	after 5 weeks
LDT	LDT 1	LDT 2	LDT 3	LDT 4
GFT	GFT 1	GFT 2	GFT 3	GFT 4
number of target word	15	15+15= 30	15+15+15= 45	45
word list	A	A + B	A + B + C	A + B + C

4.2.8 Ethical procedures

The information about the ethical issues is identical to that described in the pilot study methodology, see section 3.2.10.

4.3 Summary of Adjustments Based on the Pilot Study

The methodology for the current study was supported by pilot study observations and results, and by evidence and suggestions from previous research. The RMCG showed a slightly better performance than the FDCG in the pilot study. Therefore, the RM word cards were assigned for the word card group in the main study. The number of target words and the duration of learning was appropriate for each participant in the pilot study. Therefore, participants in the main study received the same number of target words (i.e., 15) in the same amount of time (i.e., five days a week).

To control for variability in language proficiency, vocabulary size knowledge, and WM capacity of the participants, the three groups grammar test, VST, and digit span test results were analysed prior to the learning. Also prior to learning, the three groups responded to a questionnaire relating to their language contact in order to review the participants English use in everyday life.

There were indications of how the RMCG and the GRG would perform in the main study in light of the pilot study results, and therefore insights from the pilot study were considered in

the main study design. The pilot study was a rehearsal of how to deal with the procedure of the main study from selecting participants, grouping and testing them to collecting and analysing the data. Overall, the duration of the current study, and the repetition of the use of the LDT with different length and design, allowed the researcher to overcome some of the errors that occurred in the LDT.

The design of the main study was constructed in order to respond to the research questions. In order to investigate which mode of learning resulted in superior retention during and after 15 days of learning, and in order to investigate if Arab ESL learners' performance decrease when large quantities of vocabulary input were maintained over an extended and consistent period of time, three LDTs and three GFTs were designed for the current study. The target word characteristics, such as length, phonological aspects, and derivational ending frequency were analysed before and after the participants received or were tested on them. To investigate the speed of response in the weekly immediate post-tests and the delayed post-test the LDT RT was included in the LDT design.

To respond to how well Arab ESL learners retained the target items using the Leitner box with five sections, the material for the *Leitner system* was designed and prepared, and the material for recording the movement was designed and sampled, before the pilot study and the main study took place. As for the graded readers, the design of the texts considered the number of occurrences of each target word and the number of repetitions before the pilot and the main study took place.

5 Main Study Results

5.1 Overview

The aim of the current study was to explore the approximate number of repetitions and the degree of spacing necessary for intentional learning, which might occur from employing word cards and the *Leitner system*, and compare this with incidental vocabulary learning, which might occur from using graded readers. The current study also aimed to investigate word characteristics that might affect ESL learners in general, and that might affect Arab ESL learners specifically, and it aimed to investigate whether or not the different lists were equally difficult. Finally, this study aimed to investigate the effect of the variables mentioned above on the retention of newly learned vocabulary two weeks after completion of the intervention. To investigate these aims, comparisons are made of the RMCG performance with the GRG performance over the duration of three weeks of learning, and with the performance of the CG who had no vocabulary learning guidance.

The independent variables for the current study were: the learning task (RM word cards versus graded readers versus no vocabulary guidance); word characteristics; the number of repetitions of the 45 target words (one to eight times for the RMCG, and 10 times for the GRG); and the spacing between the occurrence of the target words, i.e., from one or more days, depending on the participant performance and word difficulty, for the RMCG, and twice a week for the GRG. The spacing for the RMCG differed from that for the GRG because the RMCG participants had decided to move word cards forward or backward at any time they want. However, the GRG was introduced to target words twice a week, five times each time, and they were not given the choice to space the words, as are the RMCG. The dependent variables include the number of target vocabulary items recognised in four LDTs and produced in four GFTs. This study was designed to investigate the effect of these independent variables on the dependent variables.

As explained in the Methodology Chapter, the three groups in this study were in the second year of their Bachelor's degree. To investigate if their language proficiency levels were similar, they were required to take the VST, and their scores from a previous grammar test given by the Department of foreign languages at Albaha University were evaluated. The students' vocabulary sizes range from 3000 to 4000 words, according to Nation and Beglar's

(2007) VST, and the scores obtained on the grammar test range from 30 to 50 out of 60. Also, prior to giving the participants the tasks, the groups were asked to complete a questionnaire about their learning history and to carry out a Digit Span test (see the Methodology chapter).

In the current study, the participants took three weekly immediate post-tests (one at the end of each week), and one delayed post-test two weeks after learning, in order to test their receptive knowledge of word forms and their productive word knowledge of the form and meaning of the 45 target words. Receptive skills were measured with the LDT and productive knowledge was measured by the GFT. These tests were given to the participants after each of the three weeks of learning. Finally, participants took the LDT 4 and the GFT 4 as post-tests to test their retention of the 45 target words two weeks after the completion of the instruction program (see the Methodology chapter). The following is an evaluation of the research questions, presented in the Literature Review Chapter, which provides a quantitative and qualitative analysis of participants' performance during the learning process in addition to their scores on the three weekly immediate post-tests and the delayed post-test.

5.2 Evaluation of the Main Study Research Questions

In the following sections, the results are addressed in relation to each of the research questions. Firstly, to respond to the first research question, within group comparisons for the LDT AS and scores in the GFT were analysed for the RMCG, the GRG, and the CG in the weekly immediate post-tests and the delayed post-test. In addition, within group comparisons for the LDT RT for each group at the three time points were analysed. Following this, a between groups comparison for the LDT AS, the LDT RT, and the GFT scores was made. Furthermore, to assess the effectiveness of the *Leitner system*, the FWMs and the BWMs for the RMCG were analysed. To assess the efficiency of the graded readers, the analysis covers the scores on the graded readers CQs during fifteen days of learning.

Also, a comparison of the effect of the learning burden of different words on the three groups performance was analysed. In a qualitative analysis, the effect of word characteristics on word acquisition for both the RMCG and the GRG were analysed. The word characteristics that might cause difficulty for ESL learners were analysed first, followed by the word characteristics that might affect Arab ESL learners specifically.

Spacing between target words and the timing of the presentation of target words for the RMCG and the GRG were analysed. More specifically, the RMCG card movements in the *Leitner system,* and the GRG time of exposure to target words during the five days of learning

(i.e., beginning, middle, and end) were analysed. The following is the evaluation of the first research question:

5.2.1 Which mode of learning results in superior retention during and after fifteen days of learning?

The evaluation of this research question gives information about target word retention by the RMCG, the GRG, and the CG. To investigate the number of target words retained by both experimental groups, analyses are made concerning the recognition of the target word, the RT of response to the target word in the LDTs, and the production of the target word in the GFTs, and compared with the performance of the CG

For each of the weekly immediate post-tests and the delayed post-test, LDTs AS and LDT RTs were recorded. In addition, four GFTs were also employed to measure word production in context. Table 5.1 shows the Cronbach's alpha value of the LDT AS for tests 1, 2, and 3, and the post-tests.

Table 5.1 *Cronbach's alpha of LDT 1 AS, LDT 2 AS, LDT 3 AS and LDT 4 AS.*

LDT	**LDT 1**	**LDT 2**	**LDT 3**	**LDT 4**
Cronbach's alpha	.863	.855	.867	.789

As for the GFT scores, Cronbach's alpha values for the first, second, third, and fourth GFT were reported in Table 5.2. The Cronbach's alpha values for the LDT AS and GFT scores in the weekly tests and post-test was .7 or more, this indicates that both tests were reliable.

Table 5.2 *Cronbach's alpha of GFT1, GFT 2, GFT3, and GFT 4*

GFT	**GFT 1**	**GFT 2**	**GFT 3**	**GFT 4**
Cronbach's alpha	.926	.958	.960	.943

The following section provides information about LDT AS and GFT scores in the weekly immediate post-tests and the delayed post-test for the RMCG, GRG, and CG.

5.2.1.1 Which of the two groups (the RMCG or the GRG) learn and retain more target-word as shown by the weekly immediate post-tests and the delayed post-tests?

To respond to this question an analysis of the three groups LDTs AS and the GFT scores in the weekly immediate post tests and in the delayed post-test is presented.

Week 1 post-test

To test the participants' knowledge of the form of 15 target words, they received the first LDT. Table 5.3 shows that the RMCG had the highest LDT AS (13.12), the GRG came next (12.48), and the CG last (5.04).

Table 5.3 *Means and standard deviations of LDT 1 AS (max = 15) for the RMCG, GRG and CG*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	13.12	12.48	5.04
SD	1.01	1.19	2.70

The distribution of AS for the RMCG differed significantly from the normal distribution, according to the Kolmogorov-Smirnov Z test ($z = .213$; $p < .0005$). The AS of the GRG was also found to be not normally distributed, according to the Kolmogorov-Smirnov Z test ($z = .213$; $p<0.005$). However, the AS of the CG distribution was normally distributed, where $z = .256$ and $p = .068$. Given that the distribution of the majority of the conditions in this study was not normally distributed, it was decided to use non-parametric tests throughout the analyses reported in this chapter.

The Kruskal-Wallis H test revealed a statistically significant difference between groups (χ^2 (2) = 52.247, p < .0001). Pair-wise comparison using the Mann-Whitney test revealed a significant difference between the AS of the RMCG and the GRG, (U = 205.00, z = -2.152, p = .031, r = -.30).

The differences between the first AS of the RMCG and the CG and those of the GRG and the CG were significant, where (U = .000, z = -6.105, p <.0005, r = -.86); and (U = .000, z = -6.099, p < .0001, r = -.86), respectively. Both experimental groups therefore outperformed the CG.

The first GFT consisted of fifteen sentences in which participants were asked to complete one gap for each target word. The RMCG obtained the highest score in GFT 1, followed by the GRG, and the CG. As can be seen in Table 5.4 the RMCG obtain the highest GFT score, the GRG came next, and the CG last.

Table 5.4 *Means and standard deviations of GFT 1 scores (max = 15) for the RMCG, GRG and CG*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	11.68	8.92	.12
SD	2.11	1.57	.33

The Kruskal-Wallis H test revealed a statistically significant difference between groups (χ^2 (2) = 58.911, p < .0001). Pair-wise comparisons was made using a Mann-Whitney U test reveal a significant difference between the GFT scores of the RMCG and the GRG (U = 98.500, z = -4.197, p < .0001, r = -0.59). Furthermore, the differences between the GFT scores of the RMCG and the CG (U = .000, z = -6.353, p <.0005, r = -0.89), and those of the GRG and the CG were significant (U = .000, z = -6.360, p < .0001, r = -0.89).

Week 2 post-test

The second LDT was designed to measure participants' form recognition of 30 target words. Fifteen words were the same as the target words in the first LDT, and fifteen new target words were added. The AS was calculated out of 30 for each participant in the three groups. Table 5.5 shows the RMCG had the highest AS, the GRG came next, and the CG last.

Table 5.5 *Means and standard deviations of LDT 2 AS (max = 30) for the RMCG, GRG, and CG*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	25.40	21.64	12.28
SD	2.10	2.89	3.36

The Kruskal-Wallis H test revealed a statistically significant difference between groups (χ^2 (2) = 56.322, $p < .0001$). Pair-wise comparisons using the Mann-Whitney test revealed a significant difference between the AS of the RMCG and the GRG ($U = 421.00$, $z = -4.226$, $p < .0001$, $r = -0.59$). Furthermore, the differences between the AS of the RMCG and the CG ($U = .000$, $z = -6.075$, $p < .0001$, $r = -0.85$), and those of the GRG and the CG ($U = .9.500$, $z = -5.891$, $p < .0001$, $r = -0.83$) were significant. The two experimental groups were, therefore, again found to outperform the control group.

The second GFT consisted of 30 sentences with one word gap for each sentence. The GFT's target words were the first fifteen target words (i.e., list A) in addition to the second fifteen target words (i.e., list B). The following is an analysis of GFT 2 results. As Table 5.6 shows, the RMCG had the highest GFT scores, the GRG came next, and the CG last.

Table 5.6 *Means and standard deviations of GFT 2 scores (max = 30) for the RMCG, GRG, and CG*

group	**RMCG (N=25)**	**GRG (N=25)**	**CG (N=25)**
mean	22.36	17.16	.04
SD	3.14	3.43	.20

The Kruskal-Wallis H test revealed a statistically significant difference between groups (χ^2 (2) = 59.95, $p < .0001$). Pair-wise comparisons using the Mann-Whitney U test revealed significant differences between all groups. There was a significant difference between the GFT scores of the RMCG and the GRG ($U = 87.50$, $z = -4.381$, $p < .0001$, $r = -0.61$), between the

GFT scores of the RMCG, and the CG (U = .000, z = -6.436, p < .0001, r = -0.91), and those of the GRG and the CG (U = .000, z = -6.435, p < .0001, r = -0.91).

Week 3 post-test

The third LDT measures the participants' recognition of the 45 target words, these consisted of 30 target words from the second LDT (i.e., lists A and B) in addition to a further fifteen target words (i.e., list C). The AS was calculated out of 45 for each participant in the three groups. As seen in Table 5.7 the RMCG had the highest AS, the GRG came next, and the CG last.

Table 5.7 *Means and standard deviations of LDT 3 AS (max = 45) for the RMCG, GRG, and CG*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	35.56	30.48	17.28
SD	3.17	4.48	2.55

The Kruskal-Wallis H test revealed a statistically significant difference between groups, (χ^2 (2) = 56.39, p < .0001). Pair-wise comparisons using the Mann-Whitney U test reveal significant differences between the RMCG and the GRG (U = 87.50, z = -4.381, p < .0001 r = -0.61), between the AS scores of the RMCG and the CG (U = .000, z = -6.436, p < .0001, r = -0.91), and those of the GRG and the CG (U = .000, z = -6.435, p < .0001, r = -0.91).

In the third GFT participants complete 45 sentences with one word gap in each sentence to fill. Every gap was for a target word from lists A, B, and C. In Table 5.8 the RMCG had the highest GFT score, the GRG came next, and the CG last.

Table 5.8 *Means and standard deviations of GFT 3 scores (max = 45) for the RMCG, GRG, and CG*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	30.48	18.72	.042
SD	4.08	3.51	0.20

The Kruskal-Wallis H test revealed statistically significant differences between the groups (χ^2 (2) = 67.18, p < .0001). Pair-wise comparisons using the Mann-Whitney U test revealed a significant difference between the GFT scores of the RMCG and the GRG (U = 9.00, z = -5.907, p < .0001, r = -0.83) between the AS scores of the RMCG and the CG (U = .000, z = -6.435, p <.0005, r = -0.91), and between those of the GRG and the CG (U = .000, z = -6.444, p < .0001, r = -0.91).

Delayed post-test

The fourth LDT was designed to measure participant recognition of 45 target words two weeks after the final learning session. The target words were those from lists A, B, and C. In Table 5.9 the RMCG had the highest AS, the GRG came second, and the CG last.

Table 5.9 *Means and standard deviations of LDT 4 AS (max = 45) for the RMCG, GRG, and CG*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	34.04	31.64	20.36
SD	3.32	2.87	3.12

The Kruskal-Wallis H test revealed a statistically significant difference between the groups (χ^2 (2) = 52.575, p < .0001). Pair-wise comparisons using the Mann-Whitney U test revealed a significant difference between the AS of the RMCG and the GRG (U = 169.00, z = -2.805, p = .005, r = -0.39), the differences between the AS of the RMCG and the CG (U = 2.00, z = -6.038, p < .0001, r = -0.85), and between those of the GRG and the CG (U = 1.000, z = -6.058, p < .0001, r = -0.85).

The fourth GFT was given to the participants two weeks after the completion of the intervention. It consisted of 45 sentences with one word gap to fill in each sentence. The fourth GFT tests participant production of 45 target words (i.e., lists A, B, and C). As seen in Table 5.10 the RMCG obtained the highest GFT score, the GRG came next, and the CG last.

Table 5.10 *Means and standard deviations of GFT 4 scores (max = 45) for the RMCG, GRG, and CG*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	25.44	15.16	.04
SD	4.01	3.00	.20

The Kruskal-Wallis H test revealed a statistically significant difference between the groups (χ^2 (2) = 67.71, $p < .0001$). Pair-wise comparisons using the Mann-Whitney U test revealed a significant difference between the GFT scores of the RMCG and the GRG (U = 3.500, z = -6.006, $p < .0001$, r = - 0.84) between the GFT scores of the RMCG and the CG, (U = .000, z = -6.433, $p < .0001$, r = - 0.90), and between those of the GRG and the CG (U = .000, z = -6.437, $p < .0001$, r = - 0.91).

In sum, LDT AS and GFT scores showed that the RMCG outperformed both the GRG and the CG. In the next section a summary of participant performance is addressed with respect to the LDT AS and the GFT scores in the weekly immediate post-tests and in the delayed post-test.

1. Lexical Decision Task accuracy scores (LDT AS)

In summary, the difference between the AS of the RMCG and the GRG in LDT 2, LDT 3, and LDT 4 was significant. Furthermore, the difference between the AS of the RMCG and the CG in all four tests was significant, according to the Mann Whitney U test. Finally, the difference between the AS of the GRG and the CG in the four tests was significant, where $p <$ .0001 according to the Mann Whitney U test.

2. Gap-fill test scores (GFT scores)

In summary, the difference between the GFT scores of the RMCG and the GRG in GFT 1, GFT 2, GFT 3, and GFT 4 were significant. The CG failed to answer the GFT and their results were omitted from the GFT scores analysis.

5.2.1.2 Which of the two groups (the RMCG and the GRG) react fastest to the target-word as shown by the weekly immediate post-tests and the delayed post-tests?

The LDT RT for every correct response only was recorded. For incorrect responses, the RTs were recorded as a missing value. In the following section, the results of the RMCG, the GRG, and the CG RT of target words for each LDT in the weekly immediate post-tests and the delayed post-tests are addressed. The outliers were replaced by the mean of the RT of each group in each LDT (see Appendix 6 (A)).

Week 1 post-test

The means and standard deviations of participant RTs for correct responses to the first 15 target words were calculated for each group. These values are presented in Table 5.11.

Table 5.11 *Means and standard deviations of LDT 1 RT for the RMCG, GRG, and CG (ms)*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=24)** [1].
mean	2354.97	2756.33	2824.10
SD	256.82	265.27	949.73

The Kruskal-Wallis H test revealed a statistically significant difference between the groups (χ^2 (2) = 14.403, p = .001). Pair-wise comparison using the Mann-Whitney U test revealed a significant difference between the RMCG and the GRG (U = 91, z = -4.298, p < .0001, r = -0.60). In addition, the difference between the RTs for the RMCG and the CG was also significant, where (U = 179.00, z = -2.241, p=.025, r = -0.32).

[1] One participant in the CG was omitted from the first LDT and first GFT results because her RT in the first LDT was an outlier.

Week 2 post-test

The groups' RTs means and standard deviations are presented in Table 5.12. The results show that the RTs for all three groups were slightly different. The RMCG was the fastest group to react to the target word, the CG came next, and the GRG last.

Table 5.12 *Means and standard deviations of LDT 2 RT for the RMCG, GRG, and CG (ms)*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	2238.13	2413.40	2248.52
SD	205.03	226.17	171.98

The Kruskal-Wallis H test revealed a statistically significant difference between the groups, (χ^2 (2) = 7.526, p = .023). Pair-wise comparisons using the Mann-Whitney U revealed a significant difference between the RTs for the RMCG and the GRG, (U = 194, z = -2.299, p = .021, r = -0.32). Finally, the difference between RTs of the GRG and the CG were significant where (U = 187.000, z = -2.435, p =.015, r = -0.34).

Week 3 post-test

The CG was the fastest to react to the target word, followed by the RMCG, and the GRG. All three groups' RTs means and standard deviations are presented in Table 5.13.

Table 5.13 *Means and standard deviations of LDT 3 RT for the RMCG, GRG, and CG (ms)*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	2118.10	2277.90	1758.41
SD	263.49	236.37	230.50

The Kruskal-Wallis H reveals statistically significant differences between the groups RTs, (χ^2 (2) = 34.57, p <.0001). Pair-wise comparisons using the Mann-Whitney U revealed significant differences in RTs between the RMCG and the CG (U = 94.00, z = -4.240, p <.0001, r = -0.59), and the GRG and the CG (U = 28.00, z = -5.52, p <.0001, r = -0.78).

Delayed post-test

Table 5.14 shows the means and standard deviations for the RTs for the three groups. The GRG was the fastest, the CG came next, and RMCG last.

Table 5.14 *Means and standard deviations of LDT 4 RT for the RMCG, GRG, and CG (ms)*

group	**RMCG (n=25)**	**GRG (n=25)**	**CG (n=25)**
mean	2073.00	1785.60	1700.44
SD	152.95	154.40	205.81

The Kruskal-Wallis H test revealed no statistically significant difference between the groups (χ^2 (2) = 2.079, p = .354). Figure 5.1 summarizes the LDT RT for the three groups in the four LDTs.

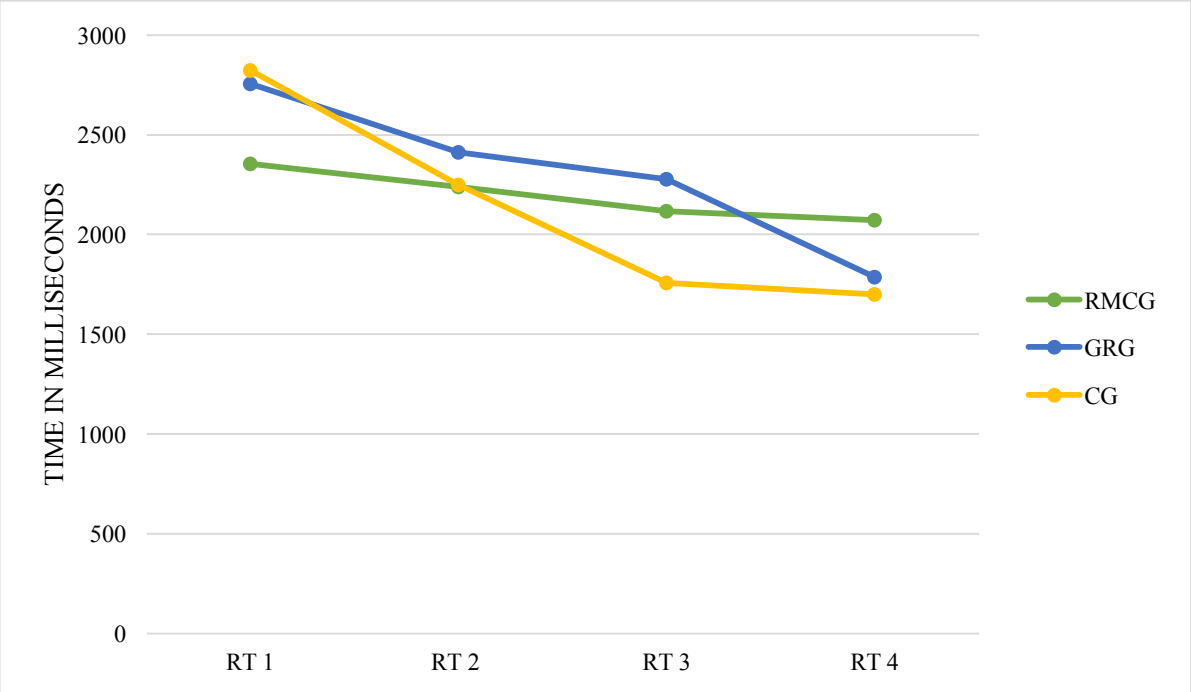


Figure 5.1 LDT RTs for RMCG, GRG, and CG in the four LDTs

In summary, the RMCG tend to respond faster than the GRG in all LDTs, except in the delayed post-test. Also, the RMCG tend to respond faster than the CG in the first and second

LDTs. Reasons for the lack of significant differences at this stage are discussed in the Discussion Chapter.

5.2.1.3 How many forward and backward moves does the RMCG make in the Leitner system?

In this section, the FWMs and the BWMs made by the RMCG in the three weeks of learning were analysed in general, followed by an analysis of the FWMs and the BWMs made each week.

At the end of each of the five learning sessions, each participant in the RMCG left their cards in the sections in which they placed them on completion of the Leitner task for the specific learning session (see the Methodology Chapter for a detailed explanation of the formula used for calculating FWMs and BWMs).

No participant moved all word cards to the fifth section before the fifth session, and all participants took the weekly immediate post-tests (i.e., LDT and the GFT) at the end of the fifth learning session each week. To calculate the mean of every movement, forward or backward, the number of movements were multiplied by the number of participants in the RMCG who execute the movement. The multiplications of the movements for each word were then summed and the summation of movement were then divided by 25 for each word (i.e., 25 is the number of RMCG participants) (see Appendix 6 (B)). The means were then used for data analysis of the FWMs and BWMs in the three weeks (see Appendix 6 (C)).

In the following section, a general explanation is given of the total frequency of FWMs and BWMs over the three weeks of learning. This is followed by an explanation of the participants' FWMs and BWMs during each individual week of learning.

The Kruskal-Wallis H test revealed no statistically significant difference in the number of FWMs between lists A, B, and C, χ^2 (2) = 3.33, p = .189. Table 5.15 shows the means and standard deviations of the total frequency of FWMs and BWMs. Clearly, the frequency of FWMs was much higher than that of the BWMs overall. The Kruskal-Wallis H test revealed no statistically significant difference in the BWMs between lists A, B, and C, χ^2 (2) = 1.983, p = .371.

Table 5.15 *Means and standard deviations of FWMs and BWMs over the three weeks of learning.*

moves	**FWMs**	**BWMs**
mean	5.19	.21
SD	.12	.13

A comparison of the RMCG participants' number of FWMs for each list independently was compared over the course of the three weeks of learning. Table 5.16 shows the means and standard deviations of FWMs of the RMCG of each list independently.

Table 5.16 *Means and standard deviations of FWMs over the three weeks of learning for lists A, B, and C independently*

list	**list A (n=15)**	**list B (n=15)**	**list C (n=15)**
mean	5.23	5.20	5.14
SD	.16	.10	.07

As for BWMs, a comparison of the RMCG participants' BWMs for each list independently over the course of three weeks of learning was made. Table 5.17 shows the BWMs of word cards in the three weeks of learning for the three lists independently.

Table 5.17 *Means and standard deviations of BWMs over the three weeks of learning for lists A, B, and C independently*

list	**list A (n=15)**	**list B (n=15)**	**list C (n=15)**
mean	.269	.195	.192
SD	.186	.102	.088

Figure 5.2 shows the FWMs means for the 45 target words over the three-week learning period in rank order from the most forward moved target word to the least. The words in list A

are the most frequently moved, followed by list B, and then list C. See Appendix 6 (C 1) for the number of FWMs for each target word.

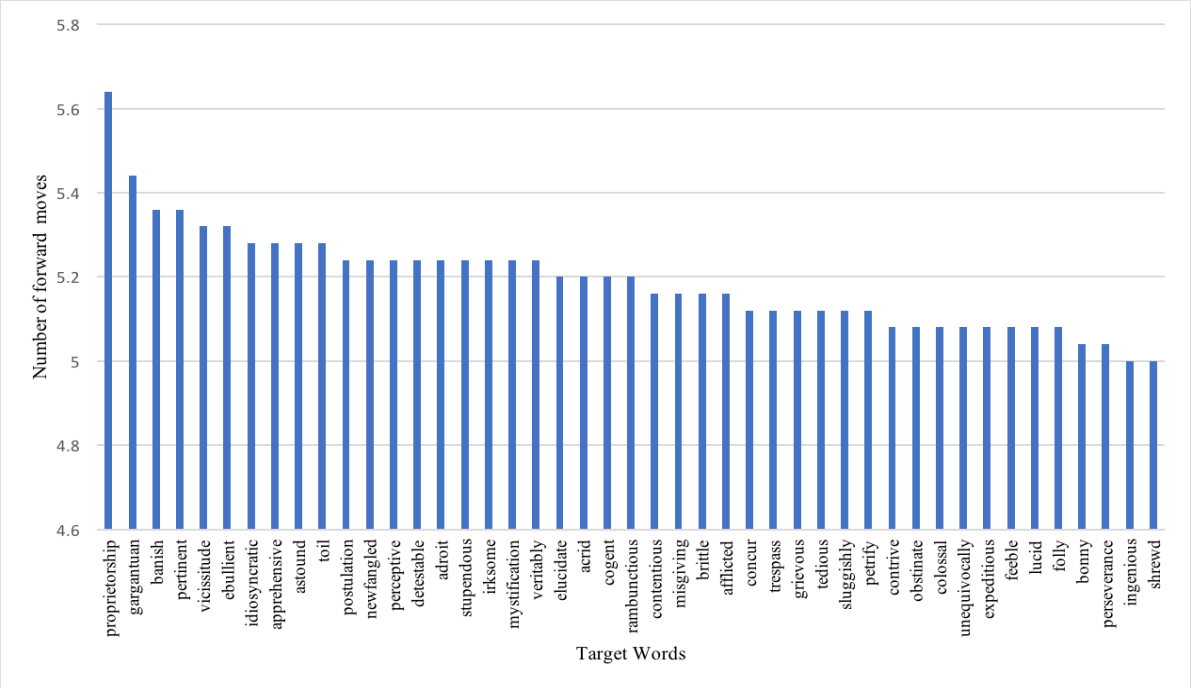


Figure 5.2 The rank order of FWMs means for each of the 45 target words over the three weeks of learning

Figure 5.3 shows the BWMs means for each of the 45 target words in rank order from the least to the greatest number of backwards moves made over the three weeks. The number of BWMs for list A was more frequent, followed by list B, and list C. Words from list A, such as *proprietorship*, *pertinent*, *banish*, *vicissitude,* and *postulation,* were the words moved backwards the most from the 45 target-word list. From list B, the words *gargantuan* and *ebullient* were most frequently moved backwards. The highest frequency of BWMs in list C was for *toil*. The words *shrewd* and *ingenious* from list A, *bonny*, *colossal,* and *unequivocally* from list B, and *feeble* and *folly* from list C showed the least frequent number of BWMs of the 45 target words. See Appendix 6 (C 2) for the number of BWMs for each target word.

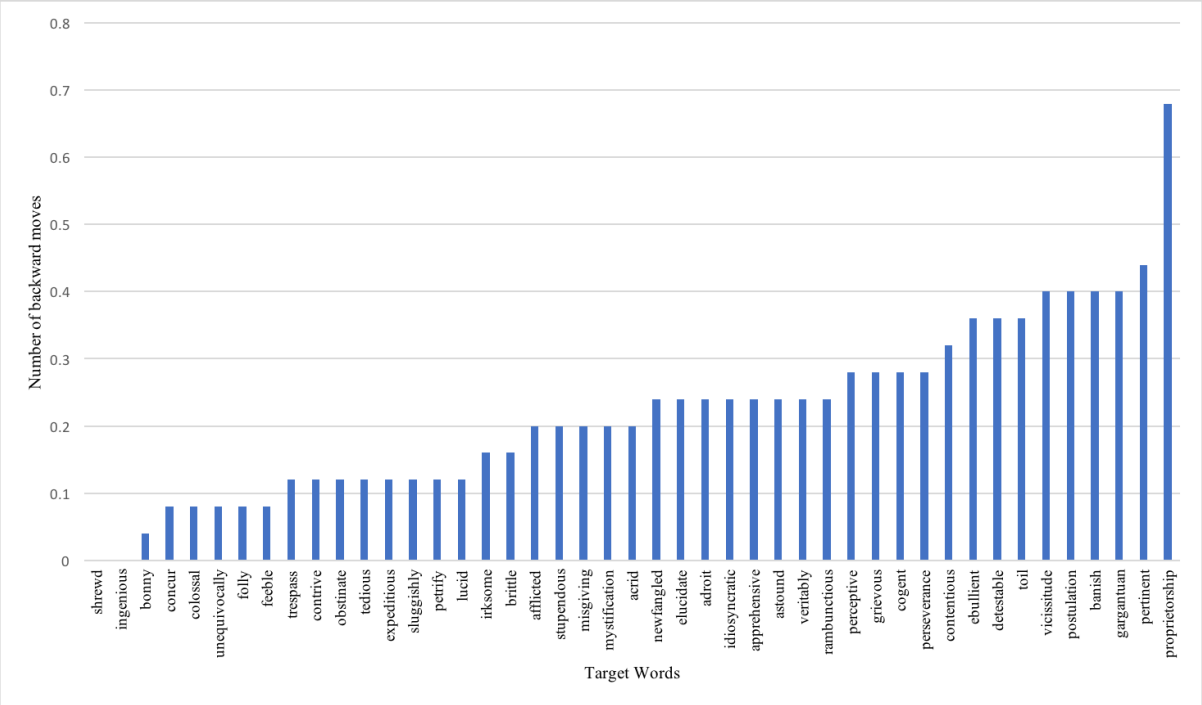


Figure 5.3 The rank order of BWMs means for the 45 target words over the three weeks

1. The RMCG performance in the Leitner system in the first week of learning

The means of FWMs for participants in the first week of learning are presented in Table 5.18.

Table 5.18 *Means and standard deviations of FWMs and BWMs in the first week of learning*

moves	**FWMs**	**BWMs**
mean	5.11	.269
SD	.152	.186

Figure 5.4 shows that FWM frequency in the first week of learning for list A varied between target words. More than 56% of the participants repeated the target words five times and only 16% of the participants made three or four FWMs. Some target word such as

proprietorship and *banish,* were repeated more than five times. See Appendix 6 (B 1) for frequency and mean of frequency of FWMs for each word in list A in the first week.

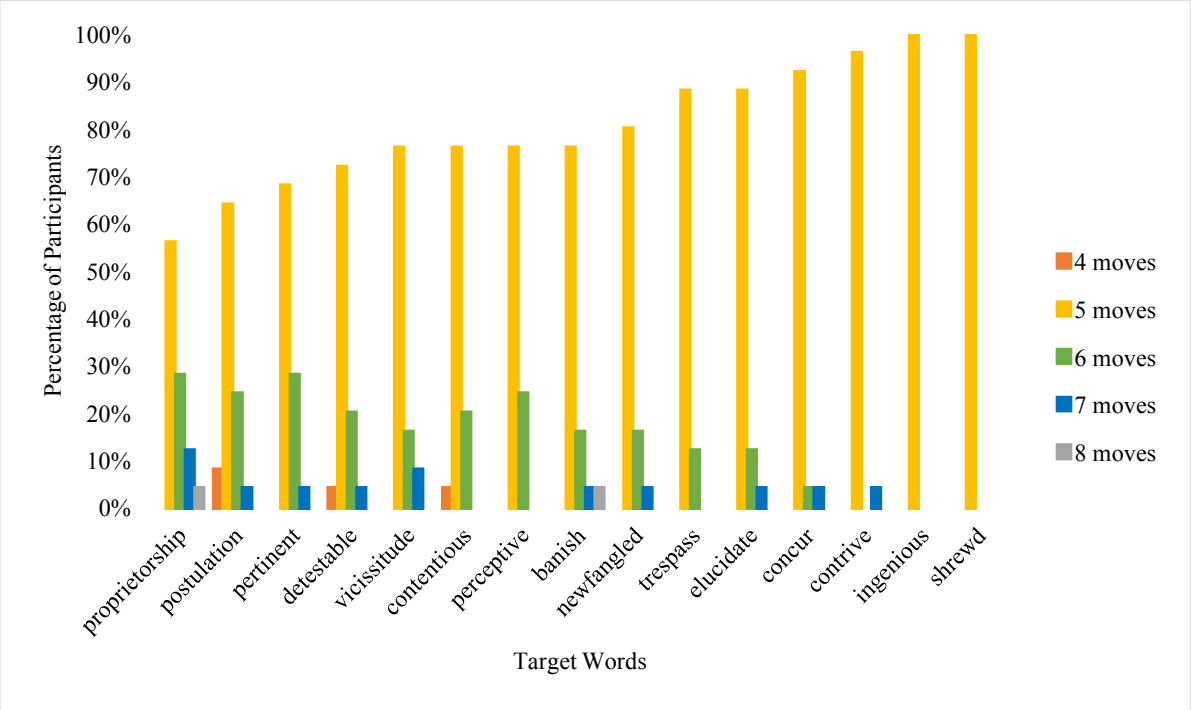


Figure 5.4 Percentage of participants' FWMs in the first week of learning

Figure 5.5 shows the BWMs for list A in the first week. Most of the participants did not make BWMs of the target words between sessions. However, some of the participants made one BWM for 13 target words, such as *contentious* and *detestable*. Twelve per cent of the participants made two BWMs for nine target words. Four per cent of the participants made three BWMs for one target word. The words *proprietorship, pertinent, vicissitude,* and *postulation* were the word cards most often moved backwards. The words *shrewd, trespass,* and *contrive* showed the lowest BWMs. See Appendix 6 (B 2) for frequency and means of BWMs of each word in list A in the first week.

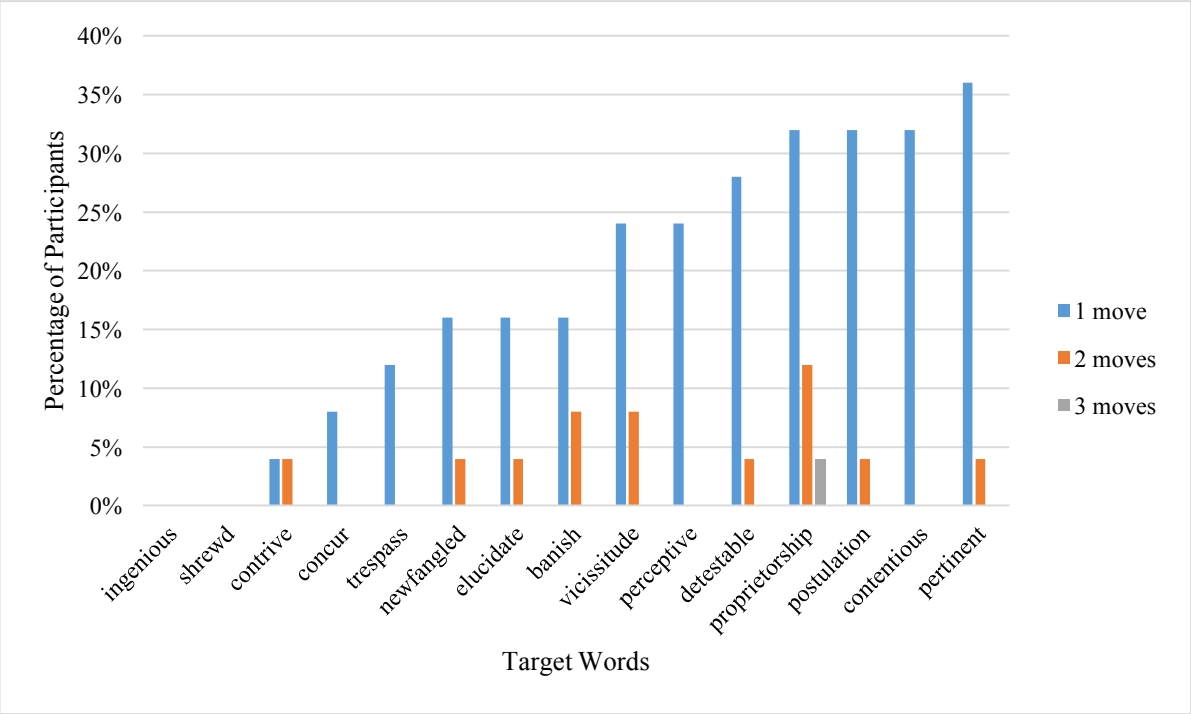


Figure 5.5 Percentage of BWMs for list A in the first week

2. The RMCG performance in the Leitner system in the second week of learning

Table 5.19 shows the mean of FWMs for list A was .09 FWMs and that there were no BWMs for list A in the second week of learning.

Table 5.19 *Means and standard deviations of FWMs and BWMs for list 'A' in the second week of learning*

moves	**FWMs**	**BWMs**
mean	.09	00
SD	.09	00

The more frequent FWMs for list A in the second week were made for the words *pertinent* (.28 FWMs) and *postulation* (.24 FWMs). As for list B, Table 5.20 shows the means of FWMs and BWMs in the second week.

Table 5.20 *Means and standard deviations of FWMs and BWMs for list 'B' in the second week of learning*

moves	**FWMs**	**BWMs**
mean	5.16	.195
SD	.103	.102

The words *gargantuan*, *stupendous*, *ebullient*, and *apprehensive* represented the most frequent FWMs in list B. Also in list B, in the second week the word *bonny* had the least frequent FWMs, followed by the words *mystification, idiosyncratic, obstinate, colossal,* and *unequivocally*. As for the BWMs for list B, the word *gargantuan* had the most frequent BWMs, followed by the word *ebullient*.

Table 5.21 shows the means and standard deviations of the FWMs and BWMs for lists A and B combined.

Table 5.21 *Means and standard deviations of FWMs and BWMs for lists 'A' and 'B' in the second week only*

moves	**FWMs**	**BWMs**
mean	2.62	.09
SD	2.58	.12

Pair-wise comparison using the Mann Whitney U test revealed a significant difference in the frequency of FWMs between lists A and B in the second week ($U = .000$, $z = -4.692$, $p < .0001$). The frequency of FWMs in the second week shows that list A differed from list B. Figure 5.6 shows the FWMs for lists A and B in the second week. Very few participants

continued to move target words from list A forward, of those few 20% moved the words *pertinent* and *postulation* forward.

List B showed a high frequency of FWMs in the second week in comparison to list A, this was because with list A many participants reached the fifth section by the end of the first week, or in the early days of the second week. As Figure 5.6 shows, the FWMs in the second week vary between target words. From 60% to 96% of the participants repeated the target words five times and about 16% of the participants made four FWMs. Some target words were repeated more than five times. For example, 32% of the participants moved the word *gargantuan* forward, 20% moved the word *adroit* forward, 16% moved the word *stupendous* forward, and 24% forwarded the word *idiosyncratic* more than five times. Overall, as in the first week for list A, five FWMs formed the pattern in the second week for list B. See Appendix 6 (B 3) for the frequency of FWMs in the second week.

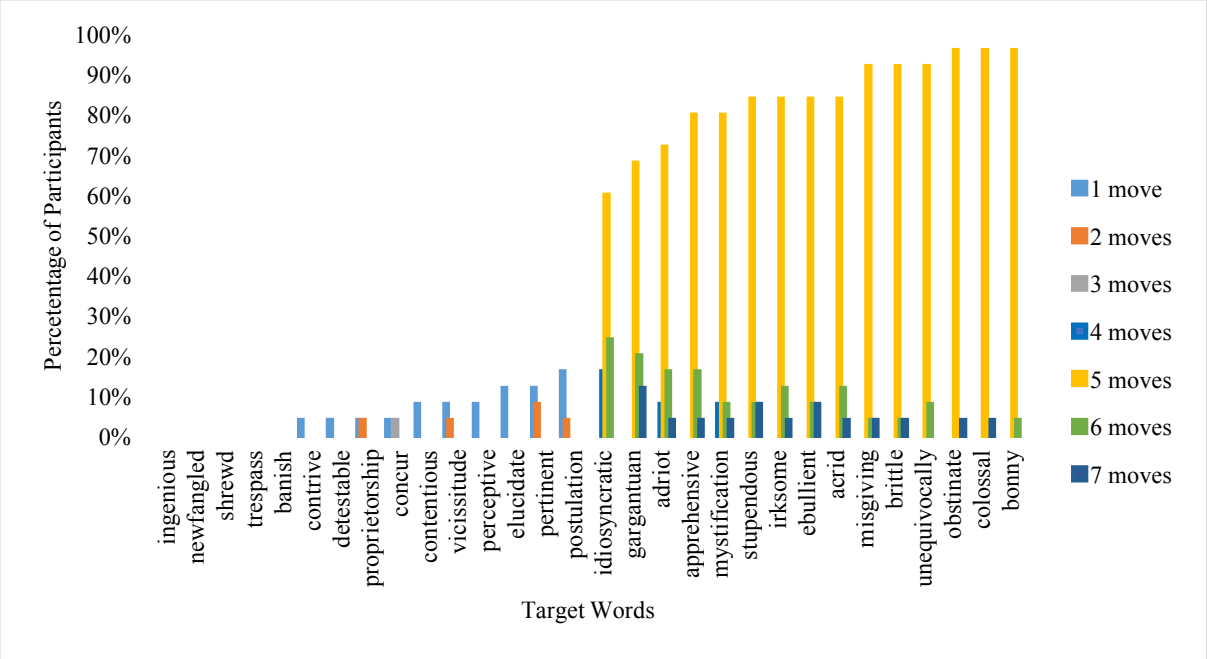


Figure 5.6 FWMs for list A and list B in the second week of learning

There were no BWMs for list A in the second week, however, words from list B were moved backward one or two times. Figure 5.7 shows the BWMs for list B in the second week. Most of the participants did not move target words backward between sessions. However, between 4% and 24% of participants made one BWM for all the 15 target words in list B. Two BWMs for 10 target words were also made by 4% to 8% of the participants in the second week.

The word *gargantuan* was moved backward once by 24% of the participants, and twice by 8% of the participants. The word *ebullient* was moved backward once by 12% of the participants, and 12% of the participants moved the word *ebullient* backwards twice. The words *unequivocally*, *obstinate*, *irksome*, and *brittle* had fewer numbers of BWMs where, respectively, 8%, 8%, 12%, and 16% of the participants moved them backward. See Appendix 6 (B 4) for the frequency of BWMs in the second week.

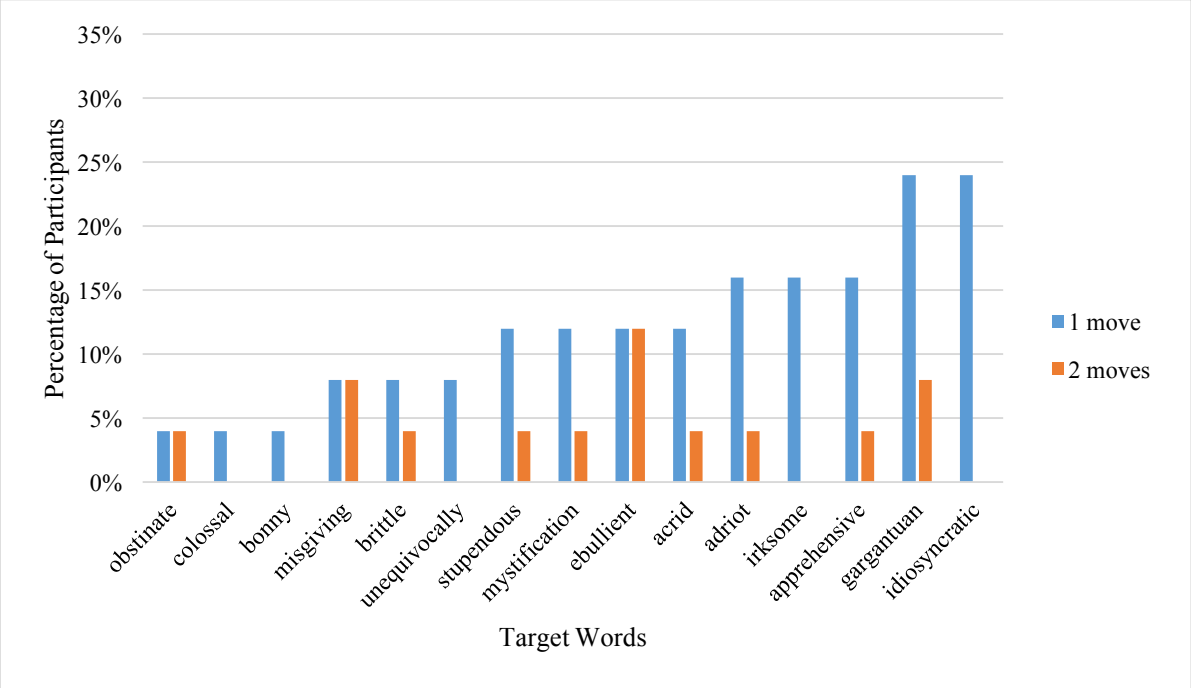


Figure 5.7 BWMs for list B in the second week of learning

3. The RMCG performance in the Leitner system in the third week of learning

The number of FWMs or BWMs for list A was zero. However, as shown in Table 5.22, the mean of FWMs for list B was .04 and there were no BWMs for list B.

Table 5.22 *Means and standard deviations of FWMs and BWMs for list B in the third week*

moves	**FWMs**	**BWMs**
mean	.04	00
SD	.05	00

The mean of FWMs for list C was 5.14 and the mean for the BWMs for list C was .19 (see Table 5.23).

Table 5.23 *Means and standard deviations of FWMs and BWMs for list C in the third week*

moves	**FWMs**	**BWMs**
mean	5.14	.19
SD	.08	.09

The mean of FWMs for lists A, B, and C combined was .02 and the mean of BWMs for lists A, B, and C combined was .06 (see Table 5.24).

Table 5.24 *Means and standard deviations of FWMs and BWMs for lists 'A', 'B' and 'C' in the third week only*

moves	**FWMs**	**BWMs**
mean	.018	.064
SD	.037	.104

Figure 5.8 shows the FWMs for lists B and C in the third week only. It shows the number of the FWMs and the percentage of participants who made these FWMs. Between 4% and 16% of the RMCG made one FWM for eight target words. The word moved forward most from list B was the word *mystification,* where 16% of the participants moved it forward once. From list B, *adroit*, *ebullient,* and *idiosyncratic* were moved forward once by 8% of the participants.

In the third week, *toil*, *astound,* and *veritably* from list C were moved forward the most. The word *toil* was moved five times by 76%, six times by 20%, and seven times by 4% of the RMCG. The word *astound* was moved forward five times by 80%, six times by 12%, and seven times by 8% of the RMCG. The word *veritably* was moved five times by 84%, six times by 12%, and eight times by 4% of the RMCG. The words *perseverance*, *expeditious,* and *feeble* showed the least frequency of FWMs. The word *perseverance* was moved forward four times

by 8%, five times by 80%, and six times by 12% of the RMCG. The words *expeditious* and *feeble* were moved forward five times by 92%, and six times by 8% of the participants. See Appendix 6 (B 5) for the FWMs for list C in the third week.

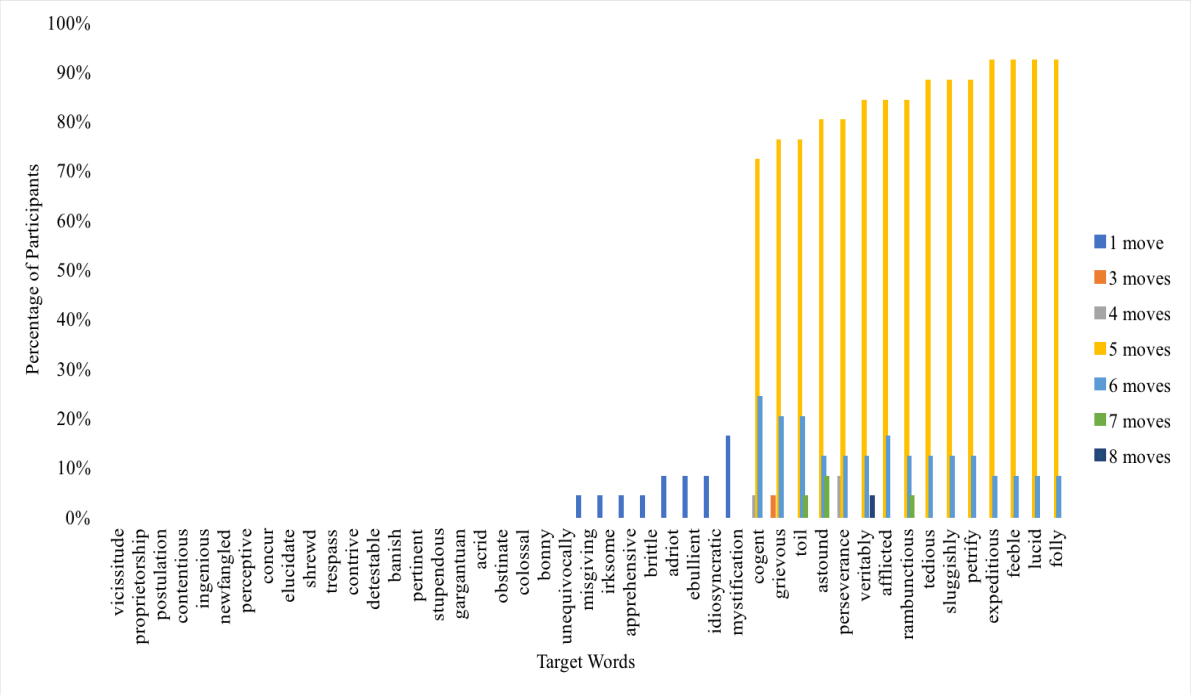


Figure 5.8 FWMs for list A, list B and list C in the third week of learning

From 12% to 28% of the RMCG moved all target words from list C backwards once. From 4% to 8% of the participants moved three target words backwards twice, only 4% moved *veritably* backwards and they did so three times. The most frequent number of BWMs made for list C were for *toil* (m = .36 BWMs), where 28% of the participants moved it backwards once, and 4% moved it backwards twice. The words *feeble* (m = 0.08 BWMs) and *folly* (m = 0.08 BWMs) show the least frequent backward moves. They were moved backwards once by 8% of the participants (see Figure 5.9). See Appendix 6 (B 6) for the BWMs for list C in the third week.

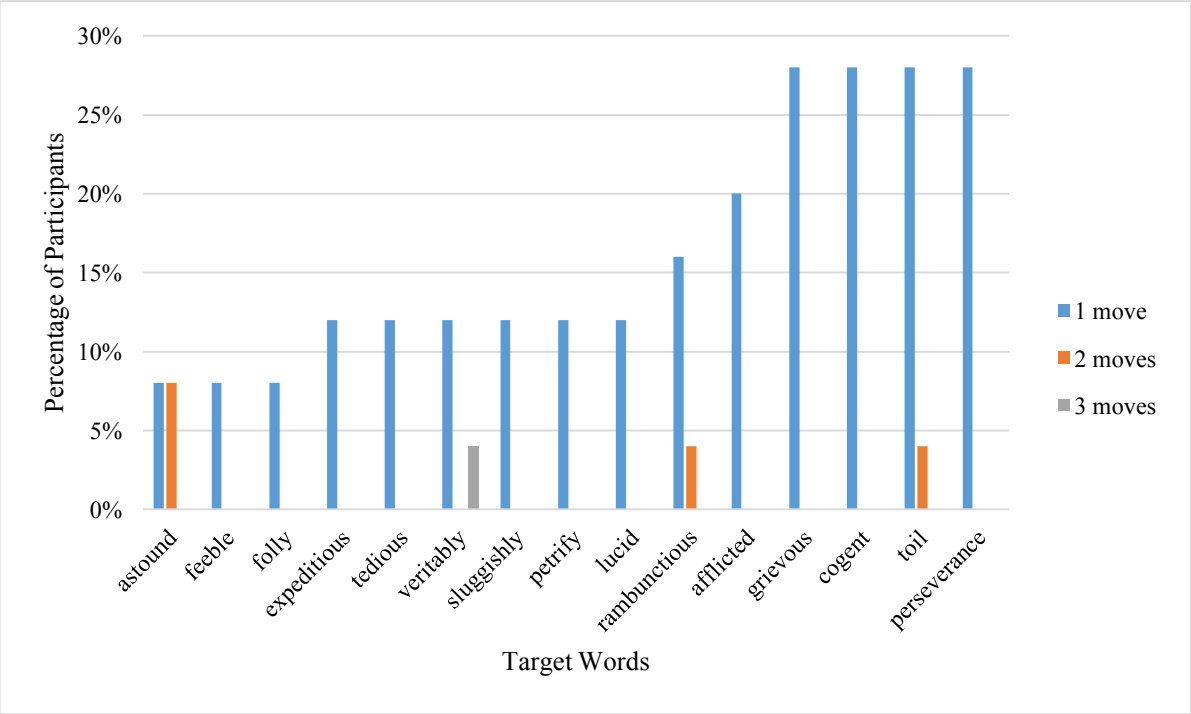


Figure 5.9 BWMs for list C in the third week of learning

5.2.1.4 How many comprehension questions does the GRG answer correctly by the end of each graded readers' texts?

To explore word retention during graded readers' learning sessions, the mean of scores for the CQs was calculated at the end of every week of learning. The total of the scores in the three weeks of learning were subsequently calculated and analysed. At the end of every graded readers' text, the participants were asked to respond to the CQs about the target word. In each week, the participants answered two CQs for every target word. The total number of CQs by the end of every week was 30 (see the Methodology Chapter).

The participants could obtain one point for every correct answer. Therefore, the maximum score was 30 at the end of every week, so in three weeks the maximum score was 90. Every point the participants gained in answering CQs means they gain points in learning target words (see the Methodology Chapter).

Overall, when the CQ scores were calculated for each participant in the GRG (i.e., maximum score of 30+30+30 = 90) the mean of the maximum scores was 68 out of 90. However, to compare the performance of the GRG for the three lists over the three weeks, the means and standard deviations for the three lists in three weeks was calculated out of 30. The

mean of the three lists was 22.93 out of 30. Table 5.25 shows the means and standard deviations for the CQs scores for each list independently.

Table 5.25 *The GRG CQs scores means and standard deviations for each list independently, the three lists together, and for the total 45 target words*

score	**list A (max = 30)**	**list B (max = 30)**	**list C (max = 30)**	**three lists (max = 30)**	**final score (max = 90)**
mean	22.84	22.88	22.08	22.93	68.80
STD	1.46	1.78	2.05	1.90	4.49

A Kruskal-Wallis H test revealed no statistically significant difference between the CQ scores for the three lists (χ^2 (2) = 11.22, p = .169).

1. CQ scores for list A in the first week of learning

The mean of the CQ scores for the first week of learning was 23.84 out of 30 (*SD* = 1.462). Figure 5.10 shows the percentage of participants who answered one question out of two, and the percentage of participants who answered both CQs. Figure 5.10 shows that more than 45% of the GRG participants were able to answer both CQs accurately for every target word. The remainder of the GRG participants answered one CQ. Sixteen per cent of the GRG participants failed to answer either of the CQs about the word *elucidate,* and 8% of the participants did not answer any of the CQs about the words *contentious, shrewd,* or *detestable*. See Appendix 6 (D 1) for the frequency of GRG CQ scores in the first week of learning.

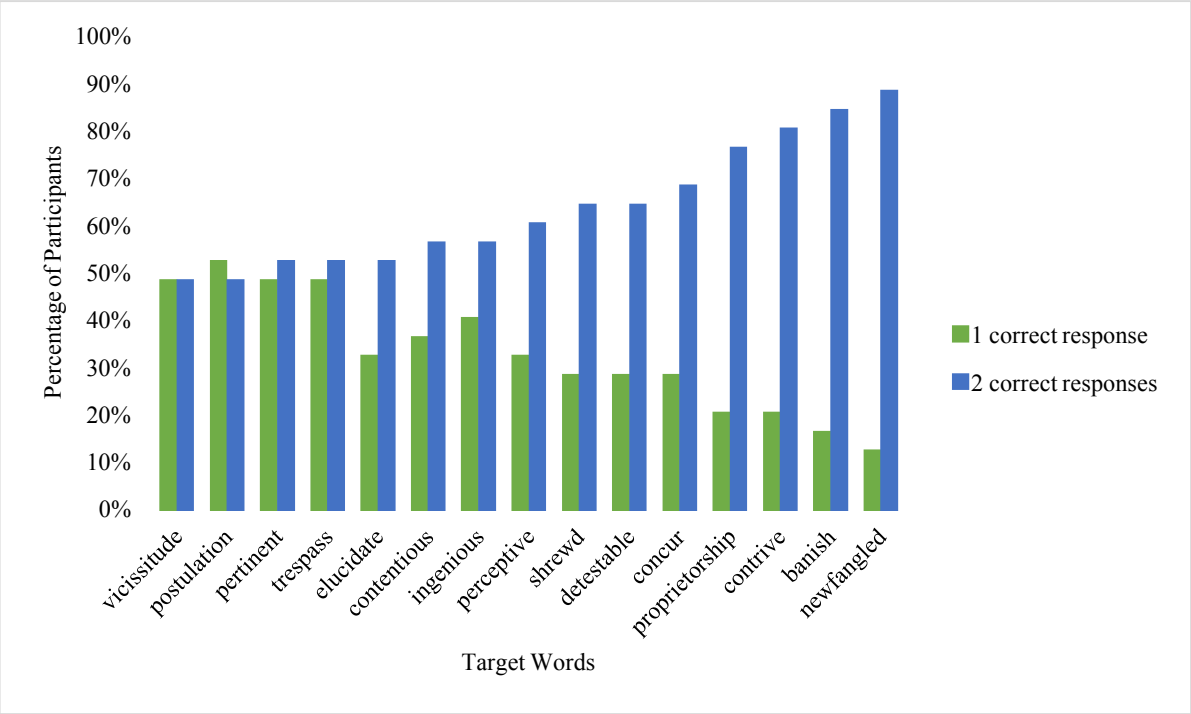


Figure 5.10 The percentage of the GRG participants who answered one CQ and the percentage of the GRG participants who answered two CQs about target words in the first week.

2. CQ scores for list B in the second week of learning

The mean of CQs scores in the second week was 22.88 out of 30 (*SD* = 1.786). Figure 5.11 shows that more than 50% of the participants were able to answer both CQs accurately for every target word. The remainder of the participants answered one CQ about the target words. As with the performance of the GRG in the first week of learning, very few participants did not answer both CQs, thus this is not presented in Figure 5.11. Sixteen per cent of the participants did not answer any of the CQs about the word *acrid,* and 12% did not answer either of the CQs about the words *ebullient* and *obstinate.* See Appendix 6 (D 2) for the frequency of participants' answers for list B in the second week.

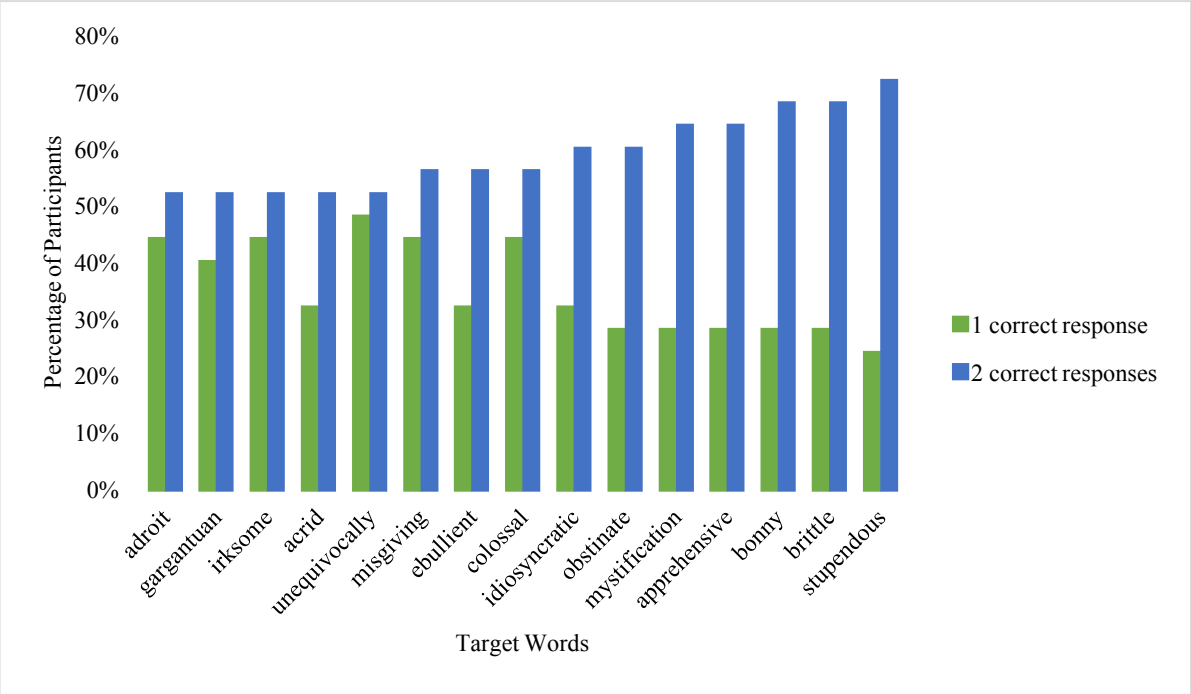


Figure 5.11 The percentage of the GRG participants who answered one CQ and the percentage of the GRG participants who answered two CQs about target words in the second week

3. CQ scores for List C in the third week of learning

The GRG average of CQ scores in the third week is 22.08 out of 30 (*SD* = 2.059). Figure 5.12 shows that more than 40% of the participants were able to answer both CQs accurately for every target word in list C. The remainder of the participants answered one CQ about the target words in list C. A very small percentage of participants failed to answer any of the CQs in the third week. Twenty per cent of the participants did not answer any of the CQs for the word *afflicted,* and 12% of the participants did not answer any of the CQs about the words *sluggishly* or *perseverance* (see Appendix 6 (D 3)).

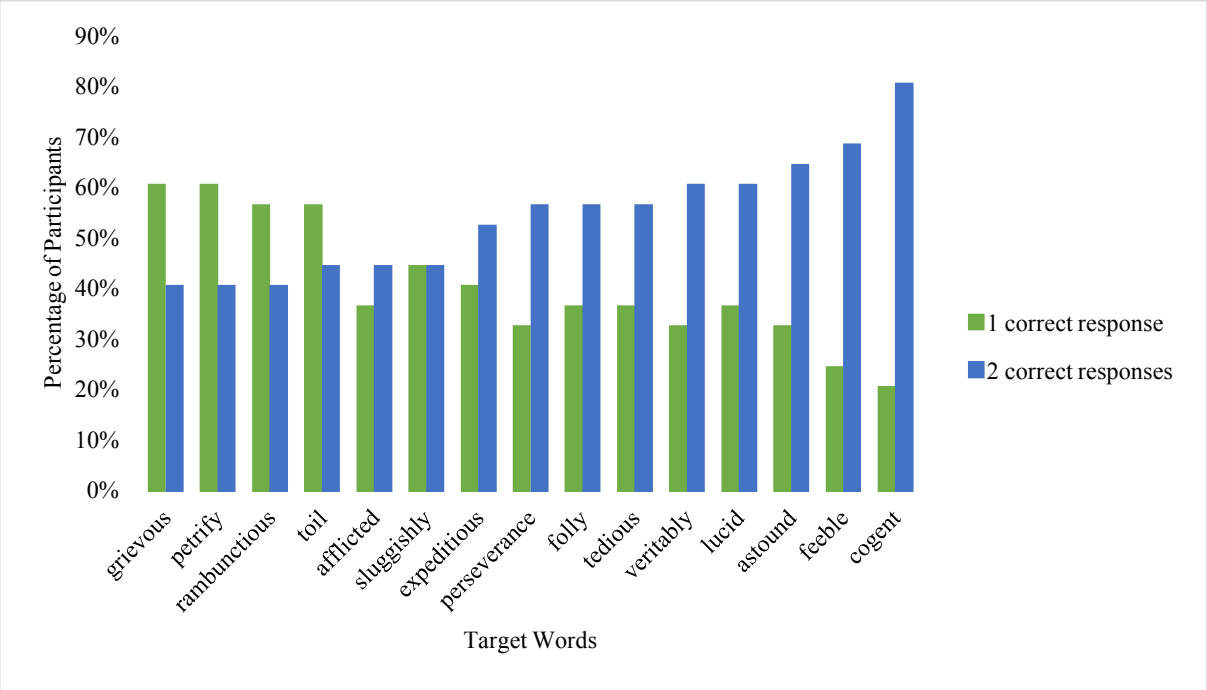


Figure 5.12 The percentage of the GRG participants who answered one CQ and the percentage of the GRG participants who answered two CQs about target words in the third week

5.2.1.5 Do Arab ESL learners' performance decrease when large quantities of vocabulary input are maintained over an extended and consistent period?

To respond to this question, a broad comparison of the weekly immediate post-tests was made to look at the overall performance of the three groups. First, a within groups comparison was made for the three groups in terms of their LDT AS, LDT RT, and the GFT scores. This was followed by a comparison between the three groups performances with respect to the LDT AS, LDT RT, and the GFT scores in all three weeks. Furthermore, a comparison of the three groups performances for lists A, B, and C in the weekly immediate post-tests and in the delayed post-test was made.

1. LDT accuracy score (AS)

a. LDT AS of the RMCG

To compare the AS in the weekly immediate post-tests and the delayed post-test, the scores were converted to percentages. As seen in Figure 5.13 and in Table 5.26, the RMCG performance declined gradually by time and the least LDT AS was in the delayed post-test.

Table 5.26 *Means of percentages and standard deviations of AS of the RMCG in the weekly tests and the post-test*

LDT AS	**AS1(max = 15)**	**AS2 (max = 30)**	**AS3 (max = 45)**	**AS4 (max = 45)**
score	13.05	25.50	35.55	33.75
percentage	87%	85%	79%	75%
SD	6.75	7.00	7.05	7.38

A Kruskal-Wallis test revealed a statistically significant difference between the AS of the RMCG in the weekly immediate post-tests and the delayed post-test (χ^2 (3) = 30.443, $p <$ .0001). A Mann Whitney U test, used for pair-wise comparisons between the AS of the RMCG, revealed that the difference between AS 2 and AS 3 (U = 178.00, z = -2.619, p = .009, r = -0.18), and the difference between AS 1 and AS 4 (U = 73.00, z = -4.690, $p <$ .0001, r = -0.33)

were significant for the RMCG. However, there were no statistically significant differences between AS 1 and AS 2 (U = 253.00, z = -1.165, p = .244), or between AS 3 and AS 4 (U = 227.50, z = -1.653, p = .098) for the RMCG.

b. LDT AS of the GRG

Figure 5.13 and Table 5.27 show that the GRG performance also declined gradually over time and the lowest AS occurred in LDT 3. The GRG performance increased in the delayed post-test in comparison to the previous LDT (i.e., LDT 3).

Table 5.27 *Means of percentages and standard deviations of AS of the GRG in the weekly immediate post-tests and the delayed post-test*

LDT AS	**AS1(max = 15)**	**AS2 (max =30)**	**AS3 (max = 45)**	**AS4 (max = 45)**
scores	12.45	21.60	30.60	31.50
percentage	83%	72%	68%	70%
SD	7.96	9.66	9.96	6.38

A Kruskal-Wallis H test revealed a statistically significant difference in the GRG performance in AS between tests (χ^2 (3) =32.435, p< .0001). A Mann Whitney U, used for pair-wise comparisons between the AS of the GRG, revealed that the difference between the AS 1 and AS 2 (U = 125.00, z = -3.672, p < .0001, r = -0.25) and the difference between AS 1 and AS 4 (U = 52.50, z = -5.08, p < .0001, r = -0.35) were significant for the GRG. However, there was no statistically significant difference between AS 2 and AS 3 (U = 229.50, z = -1.615, p = .106), or between AS 3 and AS 4 (U = 254.50, z = - 1.130, p = .253) for the GRG.

c. LDT AS of the CG

Figure 5.13 and Table 5.28 show that the AS of the CG fluctuated over time. The CG performances increased from LDT 1 to LDT 2, but decreased from LDT 3 to LDT 4, and then increased again from LDT 3 to LDT 4.

Table 5.28 *Means of percentages and standard deviations of AS of the CG in the weekly tests and the post-test*

LDT AS	**AS 1(max = 15)**	**AS 2 (max =30)**	**AS 3 (max = 45)**	**AS 4 (max = 45)**
score	5.10	12.30	17.10	20.25
percentage	34%	41%	38%	45%
SD	18.00	11.20	5.68	6.93

A Kruskal-Wallis H test revealed that there was a statistically significant difference between the CG weekly immediate post-test scores and delayed post-test scores (χ^2 (3) = 12.52, p = .006). A Mann Whitney U test, used for pair-wise comparisons between the AS for the CG, revealed that the difference between AS 3 and AS 4 (U = 138.00, z = -3.408, p = .001, r = -0.24), and between the AS 1 and AS 4 (U = 177.00, z = -2.640, p = .008, r = -0.18) for the CG were significant. This analysis also revealed that the differences between AS 1 and AS 2 (U = 229.50, z = -1.616, p = .106), and between AS 2 and AS 3 (U = 260.50, z = -1.012, p = .311) for the CG were not statistically significant.

The decrease in the percentage of word recognition from LDT 1 to LDT 2, and from the LDT 2 to the LDT 3 was 2% and 6% for the RMCG and 11% and 4% for the GRG, respectively. The percentage of the total LDT AS was calculated where the first LDT AS was out of 25, the second LDT AS was out of 30, and the third and fourth were out of 45. Figure 5.13 shows the comparison between the groups LDT AS in the weekly immediate post- tests and in the delayed post-test.

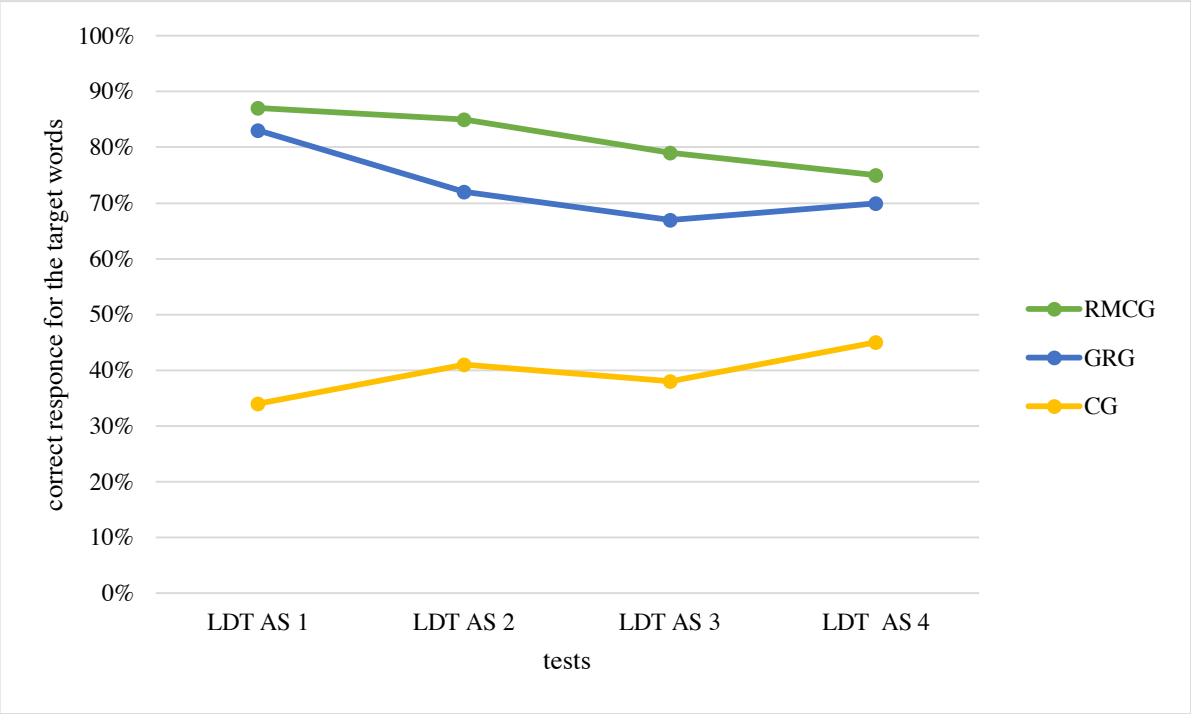


Figure 5.13 LDT percentage of AS for the RMCG, the GRG and the CG

2. Lexical Decision Task reaction time (RT)

a. The LDT RT of the RMCG

There was a noticeable decrease in the RTs from test one to the post-test for the RMCG. The RT of the RMCG gradually decreased from LDT 1 to LDT 4. The difference between the RT mean in LDT 1 and the RT mean in LDT 2 for the RMCG was 116.85 *ms*, and the difference between the RT mean in LDT 2 and the RT mean in LDT 3 was 120.02 *ms*. The RT mean in LDT 3 was 45.10 *ms* slower than the RT mean in LDT 4. Figure 5.1 and Table 5.29 show how the RT gradually decreased over time.

Table 5.29 *Means and standard deviations of RT of the RMCG in the weekly immediate post-tests and the delayed post-test (ms)*

LDT RT	**RT 1**	**RT 2**	**RT 3**	**RT 4**
mean	2354.97	2238.13	2118.10	2073.00
SD	256.82	205.03	263.49	152.95

A Kruskal-Wallis H test revealed a statistically significant difference in RT between tests for the RMCG (χ^2 (3) = 54.537, $p < .0001$). Using a Mann Whitney U test for pair-wise comparison revealed that the differences between the RT 1 and the RT 2, between the RT 2 and the RT 3, and the RT 3 and the RT 4, for the RMCG were not significant. However, the difference between RT 1 and RT 4 was statistically significant (U = 105.00, z = -4.026, $p <$ .0001).

b. The LDT RT of the GRG

The RT in LDT 2 was shown to be 342.93 *ms* faster than the RT in LDT 1. Furthermore, the RT in LDT 3 was 120.02 *ms* faster than the RT in LDT 2. Finally, the difference between RT in LDT 3 and the RT in LDT 4 was 45.10 *ms* (see Figure 5.1 and Table 5.30).

Table 5.30 *Means and standard deviation of RT of the GRG in the weekly immediate post-tests and the delayed post-test (ms)*

LDT RT	**RT 1**	**RT 2**	**RT 3**	**RT 4**
mean	2756.33	2413.40	2277.90	1785.60
SD	265.27	226.17	236.37	154.40

A Kruskal-Wallis H test was conducted to determine differences in LDT RT between the performance of the GRG in the weekly immediate post-tests and in the delayed post-test. The RTs were different for each of the four LDTs. The mean LDT RT values were significantly different between tests (χ^2 (3) = 64.540, $p < .0001$). Using a Mann Whitney U test for pair-wise comparison reveals that the difference between the RT 1 and the RT 2 (U =111.00, z = -3.910, $p < .0001$), between RT 3 and RT 4 (U =9.000, z = -5.889, $p < .0001$), and between RT 1 and RT 4 (U =.000, z = -6.063, $p < .0001$) for the GRG were significant. However, the difference between RT 2 and RT 3 was not significantly different (U = 219.00 z = -1.814, p = .070).

c. The LDT RT of the CG

The RT fluctuated from test to test. The RT 2 mean was 575.58 *ms* faster than the RT 1 mean, and the RT 3 mean was 490.109 *ms* faster than the RT 2 mean (see Figure 5.1 and Table 5.31).

Table 5.31 *Means and standard deviations of RT for the CG in the weekly immediate post-tests and the delayed post-test (ms)*

LDT RT	**RT 1**	**RT 2**	**RT 3**	**RT 4**
mean	2824.10	2248.52	1758.41	1700.44
SD	949.73	171.98	230.50	205.81

A Kruskal-Wallis H test revealed a statistically significant difference between CG tests, (χ^2 (3) = 57.547, p <.0001). Using a Mann Whitney U test for pair-wise comparison revealed that the difference between RT 1 and RT 2 (U = 158.00, z = -2.998, p = .003), between RT 2 and RT 3 (U = 25.00, z = -5.578, p <. 0001), and between RT 1 and RT 4 (U = 47.00, z = -5.151, p < .0001) were significant. However, the difference between RT 3 and RT 4 was not statistically significant, where (U = 272.00, z = -.786, p = .432).

3. Gap-fill test (GFT)

The GFT scores were converted to percentages to be able to compare tests as the differences in the maximum scores for each test were different (i.e., 15, 30, 45, and 45 points in GFT 1, GFT 2, GFT 3, and GFT 4 respectively). As seen in Figure 5.14, the RMCG and the GRG performance in GFT gradually declined from test 1 to the delayed post-test. The CG participants failed to complete the GFTs; their performance is addressed in the following sections.

a. The GFT scores of the RMCG

The first GFT scores were only 3% higher than for the second GFT for the RMCG. The second GFT scores were 7% higher than the third GFT scores. The GFT scores in the delayed post-test were 11% lower than the third GFT scores (see Table 5.32).

Table 5.32 *Means of percentages and standard deviations of GFT scores for the RMCG in the weekly immediate post- tests and the delayed post-test*

LDT AS	**GFT 1 score (max = 15)**	**GFT 2 score (max =30)**	**GFT 3 score (max = 45)**	**GFT 4 score (max = 45)**
score	11.70	22.50	30.60	25.65
percentage	78%	75%	68%	57%
SD	14.01	10.48	9.04	8.88

A Kruskal-Wallis H test revealed a statistically significant difference between the four GFT scores of the RMCG (χ^2 (3) = 37.44, $p < .0001$). Using a Mann Whitney U test for pair-wise comparison revealed that the difference between the GFT 1 scores and the GFT 2 scores for the RMCG was not significant, where (U = 260.50, z = -1.016, p = .310, r = -0.07). However, the differences between GFT 2 scores and GFT 3 scores (U = 177.00, z = -2.008, p = .008, r = -0.14), between GFT 3 scores and GFT 4 scores (U = 110.00, z = -3.941, $p < .0001$, r = -0.27), between GFT 1 scores and GFT 4 scores (U = 67.50, z = -4.767, $p < .0001$, r = -0.33) for the RMCG were significant.

b. The GFT scores of the GRG

Table 5.33 shows the means of percentages and standard deviations for the GFTs scores. There was 2% difference between GFT 1 scores and GFT 2 scores. The GFT 2 scores were 15% higher than GFT 3 scores, and GFT 3 scores were 8% higher than GFT 4 scores in the delayed post-test.

Table 5.33 *Means of percentages and standard deviations of GFT scores for the GRG in the weekly tests and the post-test*

LDT AS	**GFT 1 score (max = 15)**	**GFT 2 score (max =30)**	**GFT 3 score (max = 45)**	**GFT 4 score (max = 45)**
	8.85	17.10	18.90	15.30
percentage	59%	57%	42%	34%
SD	10.55	11.52	7.74	6.67

A Kruskal-Wallis H revealed a statistically significant difference between GFT scores of the GRG (χ^2 (3) = 65.881, p < .0001). Using a Mann Whitney U test for pair comparison revealed the difference between GFT 1 scores and GFT 2 scores was not significant, (U = 268.00, z = -.874, p = .382, r = -0.06). However, the differences between the GFT 2 scores and the GFT 3 scores (U = 64.00, z = -4.839, p <.0001, r = -0.34), between GFT 3 scores and GFT 4 scores (U = 124.50, z = -3.673, p < .0001, r = -0.25), and between GFT 1 scores and GFT 4 scores were significant (U = 6.00, z = -5.973, p < .0001, r = -0.42).

c. The GFT scores of the CG

The CG experienced a floor effect as most of the CG scores in GFT were zero or near the bottom. The GFT were too hard for the CG and they did not provide a sufficient number of responses for the four GFTs. The CG obtained .8%, .12%, .08% and .08% in GFT 1, GFT 2, GFT 3, and GFT 4, respectively. Therefore, the GFTs for the CG in the weekly immediate post-tests and the delayed post-test were not compared to the GFTs scores for the RMCG and the GRG. The decrease in the percentage of word production from GFT 1 to GFT 2, and from GFT 2 to GFT 3 were 3% and 7% for the RMCG and 2% and 15% for the GRG, respectively.

The percentage of GFT scores were calculated; the possible score for GFT 1 was 15, for GFT 2 was 30, and for GFT 3 and GFT 4 was 45. Figure 5.14 shows a comparison of the RMCG and the GRG performance in the GFTs in the weekly immediate post-test and the delayed post-test.

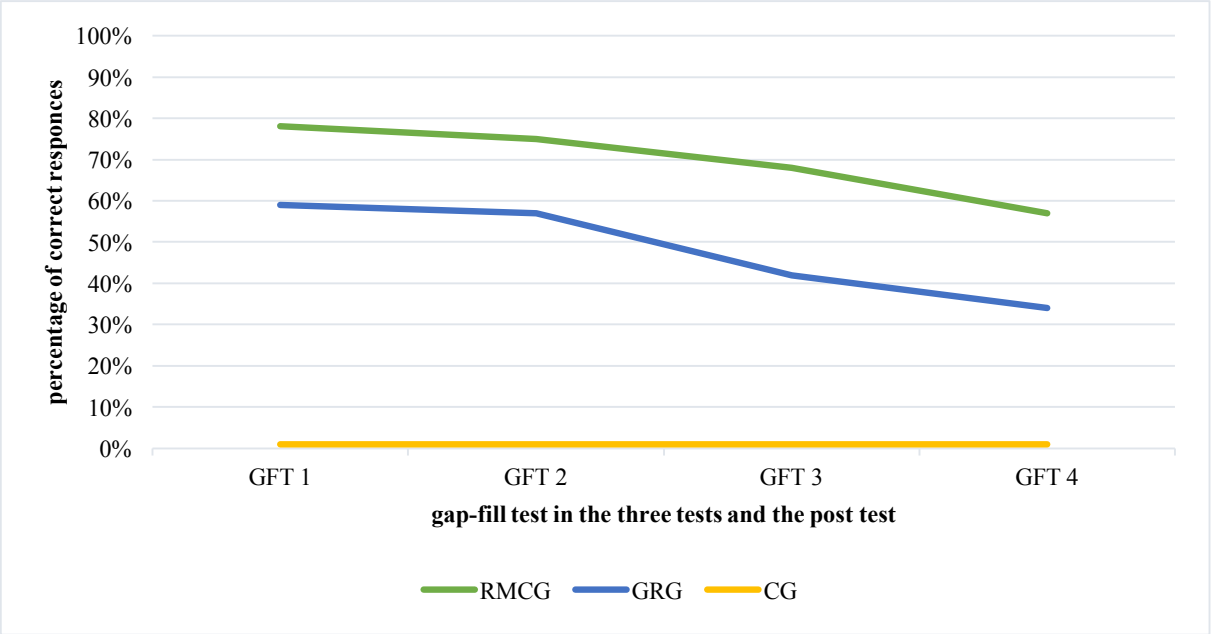


Figure 5.14 A comparison of percentages of GFT scores of three groups

Summary of significant findings

All in all, the difference between LDT 1 AS and LDT 4 AS for the RMCG was significant, where ($U = 73.00$, $z = -4.690$, $p < .0001$, $r = -0.33$). The difference between LDT 1 AS and LDT 4 AS for the GRG was significant, where ($U = 52.50$, $z = -5.089$, $p < .0001$, $r = -0.35$). The LDT AS for the CG differed significantly between LDT 3 and LDT 4, where ($U = 203.00$, $z = -2.145$, $p = .032$, $r = -0.15$). There was a significant difference in the performance of the CG in LDT AS 1 and LDT AS 4, ($U = 177.00$, $z = -2.640$, $p = .008$).

The difference between the LDT 3 RT and the LDT 4 RT for the GRG was significant, where ($U = 9.00$, $z = -5.889$, $p < .0001$). The differences between LDT 1 RT and LDT 4 RT for the RMCG ($U = 105$, $z = -4.026$, $p < .0001$), the GRG ($U = 25.00$, $z = -5.578$, $p < .0001$), and the CG ($U = 47.00$, $z = -5.151$, $p < .0001$) were significant.

The differences between GFT 3 scores and GFT 4 scores for the RMCG and the GRG were significant, where ($U = 110.00$, $z = -3.941$, $p < .0001$, $r = -0.27$) for the RMCG, and ($U = 124.50$, $z = -3.673$, $p < .000,1$ $r = -0.25$) for the GRG. The difference between GFT 1 scores and GFT 4 scores for the RMCG was significant, where ($U = 67.5$, $z = -4.767$, $p < .0001$, $r = -0.33$), and the difference between GFT 1 scores and GFT 4 scores for the GRG was significant, where ($U = 6.00$, $z = -5.973$, $p < .0001$, $r = -0.42$).

5.2.1.5.1 How does the retention of target words in lists A, B and C differ in the weekly immediate post-tests from the delayed post-test for the RMCG, the GRG, and the CG?

To investigate how participants perform for each list independently, a comparison of participant recognition, RT in the LDT, and production in the GFT for lists A, B, and C was seen in Appendix 6 (E). In summary, the three groups obtained higher AS for list A, followed by list B, then list C. Also, the RMCG and the GRG obtained better GFT scores for list A, followed by list B, then list C.

5.2.2 Do word characteristics affect word retention and learning for the RMCG and the GRG?

To respond to this research question, a qualitative analysis of the recognition and the production of the target words in relation to word characteristics that might cause difficulty for learning was made. Firstly, the qualitative analysis focuses more on the word characteristics that can cause difficulty for ESL learners in general. Following this an analysis of word characteristics that caused difficulty specifically for Arab ESL learners was made.

5.2.2.1 How do word length and derivational morphemes frequency affect word retention for Arab ESL learners?

As reported in the Methodology Chapter, the target words selected for the study had different lengths and phonological characteristics. Furthermore, target words had different derivational endings that might influence word retention (see the Methodology Chapter for an overview of the phonological and morphological characteristics of the target words). Each word was classified according to its derivational morphemes, length, and phonological characteristics (see Appendix 6 (G, H, I)).

This was an analysis of how special phonological or morphological characteristics might affect word retention. In addition, comments from the participants about some of the target words were reported. Firstly, this analysis addresses the word length characteristics and how this affects word recognition in LDTs and production in GFTs. Secondly, an analysis of morphological aspects of target words and how these might affect word recognition in LDTs and production in GFTs is considered.

1. Word length

To analyse how word length played an important role in recognition performance, each target word for the three groups was ranked from the most to the least recognised. In the following section, short (i.e., target words with one to two syllables), and easily learned target words were addressed, followed by long (i.e., target words with three to six syllables) and easily learned target words. After this, more difficult short target words were addressed, followed by long and difficult target words. The easiness and/or difficulty was determined from the average of the LDT AS and the GFT scores for the groups. Target words with a higher than average percentage of participant recognition or production were considered to be easily learned, and words with a lower than the average percentage of participant recognition or production were considered difficult to learn (see Appendix 6 (H)).

a. Short and easily learned target words

Table 5.34 shows that the most easily learned words, according to the LDT AS, were *shrewd* and *trespass. Shrewd* showed a high number in terms of RMCG and GRG recognition in the LDT, 92% of the RMCG and 80% of the GRG recognised it. Furthermore, several participants in the RMCG comment that the word *shrewd* was very easy to learn probably because of its shortness, not consisting of difficult consonant for Arab learners, and its translation ذكي in Arabic is frequently used. The second short target word with high numbers of participant recognition and production was the word *trespass.* Ninety-one per cent of the RMCG participants and 83% of the GRG recognised the word *trespass* in the LDT.

Other words with a high recognition performance for the RMCG and the GRG were *concur*, *banish,* and *bonny.* Table 5.34 shows that both groups had higher numbers of participant recognition than average for these words. Moreover, some of the GRG participants understood the meaning of the word *bonny* because it described a woman, for example, the word *bonny* occurred in the second graded reader in this context: "*I only caught a glimpse of her at the moment, but she was* ***a bonny*** *woman, with a face that a man might die for".* Some of the RMCG participants thought that this word was lovely and easy to learn, and they thought they will use it in their writing.

The RMCG differed from the GRG in the numbers of participant recognition higher than average for *contrive, folly, lucid,* and *irksome*. However, the GRG differed from the

RMCG in the numbers of participant recognition higher than average for *feeble, astound,* and *brittle.*

Finally, the CG higher than average recognition performance differed slightly from the RMCG and the GRG. *Feeble, brittle,* and *toil* did not appear in above the average numbers of participant recognition for the RMCG and the GRG. The CG had high numbers of participant recognition for these words, and also for some of the target words with high recognition performance for the RMCG and the GRG (see Table 5.34).

Table 5.34 *Most recognised short words in LDTs for the RMCG, GRG, and CG*

RMCG	No. of Syllables	percentage of participants Average 77%	GRG	No. of Syllables	Percentage of Participants Average = 70%	CG	No. of Syllables	Percentage of participants. Average =41%
shrewd	1	92%	tres.pass	2	83%	feeb.le	2	48%
tres.pass	2	91%	shrewd	1	80%	lu.cid	2	
con.cur	2	85%	feeb.le	2		ban.ish	2	
con.trive	2		bon.ny	2	76%	shrewd	1	47%
fol.ly	2	84%	ban.ish	2	75%	brit.tle	2	45%
ban.ish	2	83%	con.cur	2	74%	toil	1	44%
bon.ny	2	80%	as.tound	2		irk.some	2	
lu.cid	2		brit.tle	2	73%	bon.ny	2	
irk.some	2	79%				con.trive	2	43%
						tres.pass	2	42%

Table 5.35 also shows that the target words *shrewd* and *trespass* were the most easily learned words according to the GFT scores. A high percentage of participants in the RMCG produced the word *shrewd* in the GFT. The percentage of participant production in the GFT for the RMCG was one of the highest percentages of participant production for any of the target words. However, the number of participants who produced the word *shrewd* in the GRG (i.e., 53%) did not represent the target word with highest numbers of participant production in comparison to other words in the 45-word list, but it was higher than the average for participant production in the GRG.

Eighty-eight per cent of the RMCG produce the word *trespass,* one of the words with high numbers of participant production in the 45-word list. The word *trespass* was produced by 56% of the GRG, this was a relatively low percentage of participant production in comparison to production of other target words in the GRG. However, the word *trespass* was produced by

numbers higher than the average of participant production for the GRG. Several participants in the RMCG commented on how this word was easy to remember because of its "unique sound", i.e., it has the sound /s/ in two syllables. Several participants in the GRG noticed this word and asked about its meaning while they were reading. They were advised to guess its meaning from the context.

Table 5.35 shows *banish* and *bonny* had the highest numbers of participant production for the RMCG and the GRG. The RMCG percentages of participant production were 95% and 85% for the words *banish* and *bonny,* respectively, while the GRG percentages of participant production were 69% for both *banish* and *bonny*.

The RMCG differed from the GRG in percentages of participant production higher than average for *acrid* and *astound*. However, the GRG differed from the RMCG in percentages of participant production higher than average for *cogent* and *brittle* (see Table 5.35). The CG was not included in this section because they failed to provide considerable production scores in the GFTs.

Table 5.35 *Most production of short target words in GFTs for the RMCG and the GRG*

RMCG			**GRG**		
Target word	No. of Syllables	Percentage of participant production (Average 63%)	Target word	No. of Syllables	Percentage of participant production (Average 44%)
ban.ish	2	94%	ban.ish	2	69%
shrewd	1	89%	bon.ny	2	
tres.pass	2	88%	irk.some	2	61%
bon.ny	2	84%	tres.pass	2	56%
con.trive	2	73%	con.trive	2	54%
ac.rid	2		shrewd	1	53%
irk.some	2		brit.tle	2	
con.cur	2	67%	co.gent	2	52%
as.tound	2	66%	toil	1	48%
toil	1	64%	con.cur	2	45%

b. Long and easily learned target words

Table 5.36 shows that the RMCG had higher percentages of recognition than average of long target words than did the GRG and the CG. The most easily learned word, according to the LDT AS and GFT scores for the RMCG and the GRG, was *new-fangled.* The word *new-fangled* had the highest percentage of participant recognition in the LDT for both the RMCG,

at 92%, and for the GRG, at 85%. The RMCG responded to *new-fangled* in 2046.79 *ms,* faster than the average (see Appendix 6 (K)). The highest percentage of participant recognition for the CG was 43%, also for the word *new-fangled*. Some participants from the RMCG and the GRG reported that the word *new-fangled* was the easiest to learn because it started with the word *new,* a familiar word to them.

Table 5.36 shows that *ingenious* and *proprietorship* also gained higher than average percentages of participant recognition in the LDT for the three groups. Some participants reported that the word *proprietorship* reminded them of the word *property*. The participants were familiar with the meaning of the word *property* (i.e., *possession)* which was related to the meaning of the word *proprietorship* (i.e., *ownership*). *Ingenious* was reported as easy to learn because some of the RMCG participants associated its meaning with the word *genius,* which was close in sound and meaning. The RMCG responded to the word *ingenious* in 1831.20 *ms* and to *proprietorship* in 2103.41 *ms,* both faster than the average for the RMCG RT (see Appendix 6 (K1)). The GRG responded to the word *ingenious* in 1777.91 *ms,* faster than the average of the GRG RT (see Appendix 6 (K 2)). The CG responded slower than average to the words *ingenious* and *proprietorship* (see Appendix 6 (K 3)).

Vicissitude, *proprietorship, perceptive, contentious, obstinate, elucidate,* and *idiosyncratic* were highly recognised by more numbers than average for the RMCG and the GRG. The three groups responded to *vicissitude* and *obstinate* faster than the average RT (see Appendix 6 (K and L)). Conversely, the three groups responded to the word *perceptive* and *elucidate* more slowly than average (see Appendix 6 (K)). Although *vicissitude, contentious,* and *perceptive* were reported by some participants in the RMCG as difficult to understand because of their abstract meanings, higher than average numbers of the RMCG recognised them.

The RMCG differed from the GRG in the higher than average numbers who recognised the words *colossal, pertinent, postulations, stupendous, rambunctious, ebullient,* and *mystification.* Moreover, the RMCG responded to *colossal, stupendous,* and *mystification* faster than average. The word *stupendous* seems easy for some of the RMCG participants, some of them commented that it reminded them of the familiar word *stupid*. While the word *stupendous* means *extremely impressive*, the word *stupid* means *showing lack of intelligence*, and both words start with the sound */stju:pɪ/*. The GRG differed from the RMCG in higher than average numbers who recognised the words *sluggishly, veritably, detestable, apprehensive,* and *gargantuan* (see Table 5.36). Moreover, the GRG responded to *veritably* faster than average,

but slower than average for the rest of those words (see Appendices 6 (K, L, M)) for the target words RT).

Finally, the CG differed from the RMCG and the GRG in higher than average recognition of *perseverance*, *adroit, unequivocally,* and *tedious*. The CG responded faster than average to *unequivocally* and *perseverance* (see Appendices 6 (K, L, M) for the target words RT).

Table 5.36 *Most recognised long words in LDTs for the RMCG, GRG, and CG*

RMCG	No. of Syllables	percentage of participants Average 77%	GRG	No. of Syllables	Percentage of Participants	CG	No. of Syllables	Percentage of participants.
new.fan.gled	3	92%	new.fan.gled	3	85%	stu.pen.dous	3	53%
per.cep.tive	3	89%	in.gen.ious	3	82%	ver.i.ta.bly	4	52%
in.gen.ious	3		slug.gish.ly	3	80%	slug.gish.ly	3	50%
pro.pri.e.tor.ship	5	87%	ver.i.ta.bly	4	78%	per.se.ver.ance	4	
vi.cis.si.tude	4	86%	vi.cis.si.tude	4	76%	pro.pri.e.tor.ship	5	46%
con.ten.tious	3	85%	pro.pri.e.tor.ship	5		mys.ti.fi.ca.tion	5	
co.los.sal	3		e.lu.ci.date	4	75%	a.dr.oit	3	43%
ob.sti.nate	3	84%	per.cep.tive	3	74%	new.fan.gled	3	
per.ti.nent	3	83%	ob.sti.nate	3	73%	in.gen.ious	3	
e.lu.ci.date	4		de.test.a.ble	5	72%	un.e.quiv.o.cal.ly	6	
id.i.o.syn.crat.ic	6		id.i.o.syn.crat.ic	6		te.di.ous	3	42%
pos.tu.la.tions	4	82%	con.ten.tious	3	71%			
stu.pen.dous	3	80%	ap.pre.hen.sive	4				
ram.bunc.tious	3		gar.gan.tu.an	4				
e.bul.lient	3	79%						
mys.ti.fi.ca.tion	5							

New-fangled, proprietorship, obstinate, and *ingenious* were the most easily learned words, according to the LDT AS and GFT scores for the RMCG and the GRG. Table 5.37 shows that a higher than average percentage of participants in the RMCG and the GRG produced the words *new-fangled, proprietorship,* and *ingenious* in the GFT. *Stupendous* was highly recognised by the RMCG only, but it was highly produced by the RMCG and the GRG

(see Table 5.37). *Adroit* and *colossal* did not have higher than average recognition percentages for the RMCG and the GRG, but they had higher than average production percentages for the RMCG and the GRG (see Table 5.37).

The RMCG differed from the GRG in its higher than average production for the words *vicissitude, elucidate,* and *contentious.* Conversely, the GRG differed from the RMCG in its higher than average production of the words *tedious, sluggishly,* and *misgiving* (see Table 5.37).

Table 5.37 *Most production of long target word in GFTs for the RMCG and the GRG*

RMCG			**GRG**		
Target word	No. of Syllables	Percentage of participant production	Target word	No. of Syllables	Percentage of participant production
new.fan.gled	3	89%	te.di.ous	3	64%
vi.cis.si.tude	4	85%	stu.pen.dous	3	63%
in.gen.ious	3	84%	new.fan.gled	3	61%
pro.pri.e.tor.ship	5	77%	a.dr.oit	3	59%
co.los.sal	3	75%	in.gen.ious	3	58%
e.lu.ci.date	4	73%	slug.gish.ly	3	52%
stu.pen.dous	3	72%	pro.pri.e.tor.ship	5	51%
ob.sti.nate	3	71%	mis.giv.ing	3	49%
con.ten.tious	3	68%	co.los.sal	3	47%
a.dr.oit	3	67%			

c. Short and difficult to learn target word

Although the word *toil* is a one syllable word, it showed very low percentages of participant recognition in the LDT and production in the GFT. The RMCG showed 64% participant recognition and production of the word *toil*. The GRG showed 66% participant recognition in LDT and 48% of word production in the GFT for the word *toil*.

Both the RMCG and the GRG displayed difficulty recognising the short words *cogent, acrid,* and *grievous.* The RMCG differed from the GRG in their lower than average percentage of participant recognition of the short words *brittle, astound,* and *feeble.* Conversely, the GRG differed from the RMCG in their lower than average percentage of participant recognition of the short words *contrive* and *irksome* (see Table 5.38).

Table 5.38 Least recognised short words in LDTs for the RMCG, GRG, and CG

RMCG	**No. of Syllables**	**percentage of participants Average 77%**	**GRG**	**No. of Syllables**	**Percentage of Participants Average = 70%**	**CG**	**No. of Syllables**	**Percentage of participants. Average =41%**
ac.rid	2	76%	con.trive	2	69%	fol.ly	2	40%
brit.tle	2	74%	ac.rid	2	68%	co.gent	2	38%
as.tound	2	72%	griev.ous	2	76%	griev.ous	2	38%
griev.ous	2	68%	toil	1	66%	as.tound	2	36%
toil	1	64%	irk.some	2	65%	ac.rid	2	35%
feeb.le	2	62%	co.gent	2	60			
co.gent	2	58%						

The RMCG and the GRG seem to have difficulty producing the short words *lucid, feeble*, and *grievous.* The RMCG differed from the GRG in their lower than average numbers of participant production of the short words *folly, cogent,* and *brittle.* Conversely, the GRG differed from the RMCG in their lower than average numbers of participant production of the short words *astound* and *acrid* (see Table 5.39).

Table 5.39 *Least production of short target word in GFTs for the RMCG and the GRG*

RMCG			**GRG**		
Target word	No. of Syllables	Percentage of participant production (Average 63%)	Target word	No. of Syllables	Percentage of participant production (Average 44%)
lu.cid	2	56%	as.tound	2	40%
fol.ly	2	50%	feeb.le	2	
co.gent	2		ac.rid	2	39%
feeb.le	2		lu.cid	2	36%
griev.ous	2	48%	griev.ous	2	30%
brit.tle	2	45%			

d. Long and difficult to learn target words

Table 5.40 shows that the RMCG had less difficulty in recognising long words than did the GRG and the CG because the number of long words recognised by less than the average for the RMCG were fewer than the number of target words recognised by less than the average for the GRG and the CG. However, all three groups showed difficulty in recognising the three-syllable word *petrify*. Moreover, the RMCG and the GRG showed more difficulty in recognising the six-syllable word *unequivocally,* the four-syllable words *expeditious* and *perseverance,* and the three-syllable words *misgiving, tedious*, and *afflicted.* Lower than the RMCG average numbers of recognition was for the words *detestable, gargantuan, sluggishly, apprehensive,* and *veritably*. Conversely, lower than average numbers of the GRG recognised the words c*olossal, rambunctious adroit, mystification, pertinent, ebullient,* and *stupendous* (see Table 5.40).

Table 5.40 *Least recognised long words in LDTs for the RMCG, GRG, and CG. Percentages show recognition*

RMCG	No. of Syllables	percentage of participants	GRG	No. of Syllables	Percentage of Participants	CG	No. of Syllables	Percentage of participants
de.test.a.ble	4	76%	te.di.ous	3	68%	per.cep.tive	3	40%
ex.pe.di.tious	4		co.los.sal	3		ob.sti.nate	3	
per.se.ver.ance	4		ex.pe.di.tious	4		pet.ri.fy	3	
gar.gan.tu.an	4		un.e.quiv.o.cal.ly	6		vi.cis.si.tude	4	39%
slug.gish.ly	4		af.flict.ed	3	66%	gar.gan.tu.an	4	
mis.giv.ing	3	73%	ram.bunc.tious	3		id.i.o.syn.crat.ic	6	
ap.pre.hen.sive	4							
te.di.ous	3	72%	pet.ri.fy	3		ram.bunc.tious	3	38%
un.e.quiv.o.cal.ly	6		a.dr.oit	3	63%	de.test.a.ble	4	37%
ver.i.ta.bly	4	66%	mys.ti.fi.ca.tion	5		ap.pre.hen.sive	4	36%
pet.ri.fy	3	64%	per.ti.nent	3	62%	per.ti.nent	3	35%
af.flict.ed	3	60%	per.se.ver.ance	4		pos.tu.la.tions	4	34%
			e.bul.lient	3	61%	e.bul.lient	3	33%
			stu.pen.dous	3	57%	con.ten.tious	3	32%
			mis.giv.ing	3	53%	af.flict.ed	3	30%
						e.lu.ci.date	4	
						ex.pe.di.tious	4	

More long words seem difficult for the RMCG and the GRG to produce. Table 5.41 shows that there is a similar difficulty for both the RMCG and the GRG in producing: six-syllable words, such as *idiosyncratic;* five-syllable words, such as *mystification;* four-syllable words, such as *veritably, gargantuan,* and *perseverance;* and three-syllable words, such as *perceptive, pertinent,* and *afflicted.* The RMCG and the GRG had difficulty recognising *petrify* and also in producing it in the GFT. Finally, the RMCG differed from the GRG in their lower than average production of *tedious* and *sluggishly*. Conversely, the GRG differed from the RMCG in their lower than average production of *contentious, obstinate, unequivocally, ebullient, vicissitude,* and *elucidate* (see Table 5.41).

Table 5.41 *Least production of long target word in GFTs for the RMCG and the GRG*

RMCG			**GRG**		
Target word	No. of Syllables	Percentage of participant production (Average 63%)	Target word	No. of Syllables	Percentage of participant production (Average 44%)
ver.i.ta.bly	4	62%	con.ten.tious	3	43%
per.cep.tive	3	61%	ob.sti.nate	3	
per.ti.nent	3		un.e.quiv.o.cal.ly	6	
gar.gan.tu.an	4		per.cep.tive	3	41%
te.di.ous	3	60%	e.bul.lient	3	
slug.gish.ly	3		vi.cis.si.tude	4	
postu.la.tions	4		ver.i.ta.bly	4	36%
ap.pre.hen.sive	4		gar.gan.tu.an	4	
de.test.a.ble	5	55%	id.i.o.syn.crat.ic	6	
mys.ti.fi.ca.tion	5		de.test.a.ble	4	35%
id.i.o.syn.crat.ic	6		af.flict.ed	3	34%
pet.ri.fy	3	44%	per.ti.nent	3	32%
per.se.ver.ance	4		ex.pe.di.tious	4	
ram.bunc.tious	3	38%	ap.pre.hen.sive	4	31%
ex.pe.di.tious	4	36%	pos.tu.la.tions	4	27%
af.flict.ed	3	34%	ram.bunc.tious	3	26%
			pet.ri.fy	3	
			e.lu.ci.date	4	25%
			mys.ti.fi.ca.tion	5	21%
			per.se.ver.ance	4	20%

2. *Derivational morphemes*

Some words with no derivational morphemes might be easier to learn. One reason that might contribute to the high numbers of participant production and recognition for the word *shrewd* is because it does not consist of any derivational morpheme. However, *toil* does not consist of any derivational morpheme either, but it was difficult for the RMCG and the GRG to recognise or produce.

a. Low frequency derivational ending and easily learned words

Although *proprietorship* had one of the lowest frequent derivational morphemes, i.e., -*ship* (54 Type frequency and 2,366 Token frequency) (see Appendix 6 (G 2)), the participants might found this suffix (-*ship*) distinctive, reminding them of a '*sailing' ship*. All three groups showed higher numbers than average for participant recognition of *proprietorship.* Moreover, *proprietorship* showed higher numbers than average for the RMCG and the GRG in terms of production (see Appendix 6 (G 1)).

Although the word *obstinate* had one of the lowest frequent derivational morphemes (i.e., -*ate*, the adjective form) with 58 Type frequency and 3,645 Token frequency (see Appendix 6 (G 2)), it was one of the target words that was easily learned. All three groups displayed higher percentages than average for participant recognition of *obstinate*. Moreover, higher numbers than average of the RMCG production percentage was for *obstinate* (see Appendix 6 (G 1)).

The word *vicissitude* had one of the low frequency derivational morphemes (i.e., -*tude* 15 Type frequency and 612 Token frequency), yet it was one of the easiest words to learn (see Appendix 6 (G 2)). Higher numbers than average of the RMCG and the GRG recognition performance was for *vicissitude.* Moreover, higher numbers than average of the RMCG production was also for *vicissitude* (see Appendix 6 (G 1)).

Higher than average numbers of the RMCG recognition was for *irksome* and higher than average numbers of the RMCG and the GRG production were for *irksome* (see Appendix 6 (G 1)). *Irksome c*onsists of one of the low frequency derivational morphemes (i.e., -*some*, with 14 Type frequency and 90 Token frequency) (see Appendix 6 (G 2)), but it seems that participants were familiar with the word *some* and this might facilitated learning.

Unlike the GRG, the RMCG had higher numbers than average recognition for words with low frequency derivational morphemes: *irksome; perseverance; pertinent;* and *ebullient.* Conversely, unlike the RMCG, the GRG had higher numbers than average recognition for a word with a low frequency derivational morpheme: *gargantuan* (see Appendices 6 (G 1 and G 2)).

b. Low frequency derivational ending and more difficult words to learn

Although the GRG showed higher numbers than average for the participant production of *cogent,* it was more difficult for the three groups to recognise *cogent.* Also, although the RMCG showed higher numbers than average for participant recognition of *ebullient* and *pertinent*, it seems more difficult for the GRG to recognise these words, and more difficult for both the RMCG and the GRG to produce them. *Cogent, ebullient,* and *pertinent* contain a low

frequency derivational morpheme -*ent*_ (i.e., 105 Type frequency and 9,758 token frequency). Furthermore, it seems more difficult for the RMCG and the GRG to recognise and produce the word *petrify*. It was easier for the RMCG to recognise *perseverance,* but it was not easier to produce for the GFT (see Appendices 6 (G1 and G2)).

c. Mid frequency derivational ending and easiest words to learn

Ingenious has the mid frequency derivational morpheme -*ous* and was one of the most learned words in this study. All groups had higher numbers than average for recognition of *ingenious,* and the RMCG and the GRG showed higher numbers than average for its production. *Contentious,* with similar mid frequency derivational ending, -*ous,* shows higher numbers than average of the RMCG and the GRG for recognition, but only higher numbers than average of the RMCG for production. The RMCG and the GRG also showed higher numbers than average for recognition of *perceptive, elucidate, and idiosyncratic*. *Perceptive, elucidate,* and *idiosyncratic* contain the mid-frequency derivational morphemes *–ive, -ate,* and *–ic,* respectively (see Appendices 6 (G1 and G2)).

Unlike for the GRG, *stupendous* and *rambunctious* showed higher numbers than average of the RMCG for recognition. However, although the GRG did not show higher than average numbers of participant recognition for the word *stupendous*, it did show higher numbers than average of participant production. Both *stupendous* and *rambunctious* had the mid frequency derivational morpheme -*ous* for adjectives (i.e., 261 Type frequency and 8,116 Token frequency) (see Appendices 6 (G1 and G2)).

Colossal and *postulation* showed higher numbers than average of the RMCG for recognition. Moreover, *colossal* showed higher numbers than average of the RMCG and the GRG for production. Unlike the RMCG, the GRG showed high numbers of participant recognition of *apprehensive* and *detestable. Colossal, postulation, apprehensive,* and *detestable* contain the mid frequency derivational morphemes -*al -ation, -ive,* and *–able,* respectively (see Appendices 6 (G 1 and G 2)).

d. Mid-frequency derivational ending and more difficult words to learn

Grievous and *expeditious* seem to be the most difficult words to learn, where higher numbers than average of the three groups neither recognised nor produced them. *Grievous* and *expeditious* contain the mid frequency derivational morpheme -*ous* for adjectives (i.e., 261 Type frequency and 8,116 token frequency), but this did not facilitate learning. *Tedious,* with the same mid frequency derivational morpheme -*ous,* was not easily recognised by the RMCG

or the GRG, nor is it easily produced by the RMCG. The GRG showed difficulty in recognising most of the target words which contain the mid frequency derivational morpheme -*ous* for adjectives (i.e., 261 Type frequency and 8,116 Token frequency) (see Appendices 6 (G 1 and G 2)).

The GRG showed lower numbers than average for participant recognition of *stupendous, grievous, tedious, expeditious,* and *rambunctious.* Also, the GRG showed difficulty recognising *mystification, colossal,* and *postulation* which contain the mid frequency derivational morphemes -*ation, -al,* and –*ion,* respectively. Moreover, the GRG show difficulty in producing *mystification* and *postulation* with the mid frequency derivational morpheme -*ation,* and -*ion* respectively (see Appendix 6 (G 1)).

Higher than average numbers of the RMCG and the GRG recognised *perceptive* with the mid frequency derivational morpheme –*ive,* or *idiosyncratic* with the mid frequency derivational morpheme -*ic.* However, it seems difficult for the RMCG and the GRG to produce these words. Moreover, the GRG showed difficulty in recognising and producing *contrive* with its mid frequency derivational morpheme -*ive.* The GRG showed higher numbers than average for the groups recognition of *apprehensive*, but the RMCG did not show higher numbers than average for recognition of this word. Also, both the RMCG and the GRG found it difficult to produce *apprehensive* (see Appendices 6 (G 1)).

e. High frequency derivational ending and easiest words to learn

The word *new-fangled* has the derivational morpheme -*ed* which is highly frequent (i.e., 1,593 Type frequency and 29,633 Token frequency), according to the spoken element of the BNC. All groups showed higher than average recognition of *new-fangled,* and more than average numbers of the RMCG and the GRG produced *new-fangled.* However, the word *afflicted* has the same high frequency derivational morpheme, -*ed*, yet it showed lower than average numbers of participant recognition and production for the three groups (see Appendices 6 (G 1 and G 2)).

More than average numbers of the RMCG and the GRG easily produced *misgiving* with its high frequency derivational morpheme, -*ing.* The GRG showed higher numbers than average of participant recognition for words ending with the high frequency derivational morpheme –*ly,* such as *veritably* and *sluggishly*. Moreover, the CG showed higher numbers than average of participant recognition for all target words ending with the high frequency derivational morpheme -*ly.* The GRG easily produced *sluggishly* (see Appendices 6 (G 1)).

f. High frequency derivational ending and more difficult words to learn

Although *afflicted* has the high frequency derivational morpheme -*ed*, it showed lower numbers than average of participant recognition and production for the three groups. Unlike for the word *new-fangled,* the high frequency derivational morpheme -*ed* did not facilitate learning of the word *afflicted* (see Appendices 6 (G 1 and G 2)).

Although the RMCG and the GRG easily produced *misgiving* with its high frequency derivational morpheme *–ing*, fewer than average numbers from both groups recognised them in the LDT. Although the GRG showed higher than average participant recognition for words ending with the high frequency derivational morpheme -*ly, unequivocally* is an exceptional case. The RMCG and the GRG showed difficulty in recognising and producing *unequivocally.* Moreover, the RMCG did not recognise nor produce all the target words ending with *–ly* (see Appendices 6 (G 1 and G 2)).

5.2.2.1.1 How do phonological characteristics affect word retention for Arab ESL learners?

1. Difficult phonological aspects of the target word

Words with the sound /p/ were compared for the numbers of the RMCG, GRG, and CG in terms of recognition and production. The words *proprietorship* /prəˈpraɪətəʃɪp/ *and trespass*/ˈtrespəs/ both have the sound /p/, but this did not cause difficulty for the three groups. Appendix 6 (I) shows that the three groups displayed higher numbers than average for participant recognition of these words, and it showed higher numbers than average of the RMCG and the GRG for production of these words. Yet, *perseverance* /pɜːsɪˈvɪərəns/ and *petrify* /ˈpetrɪfaɪ/ both contain the sound /p/ and the RMCG and the GRG showed difficulty in recognising and producing them. The RMCG showed higher numbers than average of recognition for more target words with the sound /p/ than did the GRG. For example, the RMCG showed higher numbers than the average for recognition of *postulation* /ˈpɒstjuleɪeɪʃn/*, pertinent*/ˈpɜːtɪnənt/, and *stupendous* /stjuːˈpendəs/. The GRG showed higher numbers than average for participant recognition of the word *apprehensive* /ˌæprɪˈhensɪv/.

Words with the sound /v/ were compared for the numbers of the RMCG, GRG, and the CG in terms of recognition and production. *Vicissitude* /vɪˈsɪsɪtjuːd/ has the sound /v/, but this did not cause difficulty in its recognition for the RMCG or the GRG, nor in its production for the RMCG. Similarly, *contrive* /kənˈtraɪv/, which also contains the sound /v/, shows higher

numbers than average for the RMCG's recognition, and higher numbers than average for both the RMCG and the GRG in terms of production. *Perceptive* /pəˈseptɪv/ contains not only the sound /p/ but also the sound /v/, but this did not cause difficulty in recognising them for the RMCG and the GRG. However, words with the /v/ sound, like *misgiving* /ˌmɪsˈgɪvɪŋ/, *perseverance* /ˌpɜːsɪˈvɪərəns /, and *unequivocally* /ˌʌnɪˈkwɪvəkəli/, showed lower numbers than average of the RMCG and the GRG for recognition and production (see Appendix 6 (I)).

Finally, words with the sound /g/ were compared for the numbers of the RMCG, GRG, and the CG in terms of recognition and production. *New-fangled* /ˌnjuːˈfæŋgld/ has the sound /g/, but this did not cause difficulty in its recognition for the three groups, nor in its production for the RMCG and the GRG. However, *gargantuan* /gɑːˈgæntʃuən/ and *grievous* /ˈgriːvəs/, which also contain the sound /g/, showed lower percentages than average of recognition and production for the three groups. Similarly, the RMCG showed difficulty in recognising and producing *sluggishly* /ˈslʌgɪʃli/ with the sound /g/, and the GRG showed difficulty in producing it (see Appendix 6 (I)).

a. Consonant clusters

The consonant cluster /pr/ in *proprietorship* /prəˈpraɪətəʃɪp/ did not cause difficulty for the three groups in its recognition, nor for the RMCG and the GRG in its production. Conversely, the mirror shape d/ and /b/ in *detestable* /dɪˈtestəbl/ might had cause difficulty for the RMCG as it showed lower numbers than average of participant recognition and production (see Appendix 6 (I)).

There was no observation made of Arab ESL learners avoiding the difficulty of initial consonant clusters in *proprietorship* and inserting an epenthetic vowel between consonants in the initial consonant cluster in the GFT as suggested by Catford, Palmer, McCarus, Moray, and Sinder (1974).

b. List difficulty

Concur, contrive, cogent, contentious, and *colossal* have orthographic and phonological similarities, to some extent. The RMCG did not show difficulty in recognising and producing *concur, contrive, contentious,* or *colossal.* More than average numbers of the RMCG recognised and produced *concur, contrive, contentious,* and *colossal.* Furthermore, the GRG showed less difficulty in recognising and producing *concur* and *contentious.* The GRG also

showed less difficulty in producing *colossal and contrive.* All three groups showed difficulty in recognising and producing *cogent.*

Perceptive, perseverance, apprehensive, and *pertinent* have close phonological and orthographical features, as do *misgiven* and *mystification,* and *rambunctious* and *expeditious*. Both the RMCG and the GRG showed difficulty in recognising and producing these words. Some of the RMCG participants said that they were confused between the words *feeble* and *folly,* while others commented that the words *brittle* and *ebullient* were confusing to learn because they have similar orthographic and phonological characteristics. The RMCG had less difficulty in recognising *folly,* while the GRG had less difficulty in recognising and producing *feeble.* Both the RMCG and the GRG showed difficulty in producing *ebullient,* and the RMCG showed difficulty in producing *feeble* and *folly*. Furthermore, the RMCG showed less difficulty in recognising *ebullient,* while the GRG showed less difficulty in recognising and producing *brittle.* Both the RMCG and the GRG showed difficulty in producing *ebullient,* and the RMCG showed difficulty in producing *ebullient* (see Appendix 6 (F)).

c. Vowel blindness

Vowel blindness for Arab ESL learners occurs when learners have difficulty recognising short vowels (Saigh & Schmitt, 2012). In this analysis, the short vowels which exist in the target words are /ʌ/, /ɛ/, /a/, /ɪ/, /ɒ/, /i/, and /ʊ/. The long vowels found in the target words are /uː/ and /iː/. The diphthongs and the long vowels /ɑː/ and /əː/ were not used in this analysis. The diphthongs are not very similar to short vowels or the long vowels, and the long vowels /ɑː/ and /əː/ are the sound /r/ transferred to long vowels. Arab ESL learners seem to pronounce the /r/ sound in *gargantuan* rather than the long vowel /ɑː/, and pronounce the /r/ sound in *concur* rather than the long vowel /əː/ at the end (see Appendix 6 (J)).

Short vowels

Unlike the GRG, the RMCG showed higher numbers than average in recognition of target words with three or more short vowels, such as *mystification* /mɪstɪfɪˈkeɪʃən/ which contains the short vowel /ɪ/ four times, and *idiosyncratic* /ˌɪdɪəʊsɪŋˈkratɪk/ which contains the short vowel /ɪ/ three times. The RMCG showed higher numbers than average for production of the target word *veritably* /ˈvɛrɪtəbli/ which contains three short vowels, /ɛ/, /ɪ/, /ɪ/, and *vicissitud*e /vɪˈsɪsɪtjuːd/ in which the short vowel /ɪ/ occurs three times. *Elucidate* /ɪˈljuːsɪdeɪt/ contains the short vowel /i/ twice and showed high numbers of recognition and production by

the RMCG. Finally, *folly* /ˈfɒli/ contains the two short vowels /ɒ/ and /i/ and showed higher numbers than average for recognition by the RMCG (see Appendix 6 (J)).

Moreover, the RMCG showed higher numbers than average for recognition of target words: with two short vowels, such as *rambunctious* /ramˈbʌŋkʃəs/ with /ʌ/ and /a/; *ebullient* /ɪˈbʌljənt/, with the short vowels /ʌ/ and /ɪ/; and *perceptive* /pəˈsɛptɪv/, with the short vowels /ɛ/ and /ɪ/. Also, RMCG showed higher numbers than average in recognition of the target word *postulation* /pɒstjʊˈleɪʃən/, with the short vowel /ɒ/ (see Appendix 6 (J)).

Although the GRG showed more difficulty than the RMCG in recognising and producing target words with short vowels, the GRG obtained higher numbers than average in recognition and production of a target word with three short vowels. *Misgiving* /mɪsˈgɪvɪŋ/, with the occurrence of the short vowel /ɪ/ three times, was one of the target words with high GRG recognition and production. Moreover, both the RMCG and the GRG highly recognised and produced *banish* /ˈbanɪʃ/, with the two short vowels /a/ and /ɪ/, and the target word *bonny* /ˈbɒni/, with the two short vowels /ɒ/ and /i/. Also, both the RMCG and the GRG highly recognised *obstinate* /ˈɒbstɪnət/, with the two short vowels /ɒ/ and /ɪ/ (see Appendix 6 (J)).

Proprietorship /prəˈprʌɪətəʃɪp/ has the short vowel /ʌ/ and the numbers for its recognition and production among the RMCG and the GRG was higher than average for this word. Yet, for the word *sluggishly* /ˈslʌgɪʃəli/, with the short vowel /ʌ/, but also with other short vowels, /ɪ/ and /i/, the RMCG showed difficulty in recognising and producing it. *Rambunctious* /ramˈbʌŋkʃəs/ has the short vowel /ʌ/ in addition to /a/, and *ebullient* /ɪˈbʌljənt/ has the short vowel /ʌ/ in addition to /ɪ/, and they were both difficult for the GRG to recognise and for the RMCG and the GRG to produce (see Appendix 6 (J)).

Trespass /ˈtrɛspəs/ has the short vowel /ɛ/ and the numbers of recognition and production by the RMCG and the GRG were higher than average for this word. *Contentious* /kənˈtɛnʃəs/ also has the short vowel /ɛ/ and was highly recognised by the RMCG and the GRG, and produced by the RMCG. However, *perceptive* /pəˈsɛptɪv/ and *veritably* /ˈvɛrɪtəbli/, with the short vowel /ɛ/, were difficult for the RMCG to recognise and for both the RMCG and the GRG to produce. Moreover, *petrify* /ˈpɛtrɪfʌɪ/, with the short vowel /ɛ/ was one of the difficult words to recognise and produce for both the RMCG and the GRG (see Appendix 6 (J)).

Moreover, some target words with one short vowel were considered difficult for both the RMCG and the GRG, e.g., *afflicted* /əˈflɪktəd/, with the short vowel /ɪ/. However, *brittle* /ˈbrɪtəl/, which also has the short vowel /ɪ/, was highly produced by both the RMCG and the GRG (see Appendix 6 (J)).

The RMCG and the GRG might had been experiencing difficulty in learning some of the target words with more than one short vowel. The RMCG and the GRG showed lower numbers than average for participant recognition and production of *unequivocally*

/ˌʌnɪˈkwɪvəkəli/, with four short vowels /ʌ/, /ɪ/, /ɪ/, /ɪ/, and *expeditious* /ˌɛkspɪˈdɪʃəs/, with three short vowels /ɛ/, /ɪ/, /ɪ/ (see Appendix 6 (J)).

Long vowels

Many target words with the long vowel /u:/ were highly recognised and produced. *Shrewd* /ʃru:d/ was highly recognised and produced by higher numbers than average in the RMCG and the GRG. Also, *stupendous* /stju:ˈpɛndəs/ was easily recognised by the RMCG and produced by both the RMCG and the GRG. Furthermore, *vicissitude* /vɪˈsɪsɪtju:d/ was highly recognised and produced by the RMCG, and highly recognised by the GRG. *Elucidate* /ɪˈlju:sɪdeɪt/ was highly recognised and produced by the RMCG, and recognised by the GRG. However, *lucid* /ˈlu:sɪd/, with the long vowel /u:/, was difficult for the GRG to recognise and for both the RMCG and the GRG to produce (see Appendix 6 (J)).

Fewer target words with the long vowel /i:/ were highly recognised and produced. Exceptionally, *ingenious* /ɪmˈdʒi:niəs/, with the long vowel /i:/, was highly recognised and produced by the RMCG, and highly recognised by the GRG. However, *feeble* /ˈfi:bl/ and *tedious* /ˈti:dɪəs/ were not highly recognised by the RMCG nor produced by both the RMCG and the GRG. Moreover, *grievous* /ˈgri:vəs/ showed very low recognition and production rates among the RMCG and the GRG (see Appendix 6 (J)).

d. Translation

In some instances, it seems that translation might contributed to the learning of some target words. Several participants in the RMCG commented that the word *shrewd* was very easy to learn. *Shrewd* has only one word synonym in Arabic ذكي, its meaning was easy to learn from the word cards and to retrieve from the graded readers as this meaning is used frequently in Arabic. However, *contentious* was reported by some participants in the RMCG as difficult to understand because of its abstract meaning which is not so frequently used in Arabic. Yet, 85% of the RMCG recognised the word *contentious* in the LDT, and 68% produced it in the GFT.

The word *perceptive* was reported as difficult to learn for the RMCG, although it shows high numbers of participant recognition in the LDT (i.e., 89%). The reason it might had caused this feeling of difficulty was because its meaning requires a two-word translation into Arabic *(وجهة نظر)*. One of the words that the RMCG and the GRG found difficult to recognise and produce was the word *pertinent (ذا صلة)* because they said its meaning in Arabic is not

very frequently used and it has no synonym, rather it has a two-word translation (see Appendix 6 (F)).

5.2.3 To what extent does repetition and spacing affect learning using word cards and graded readers?

To respond to this question, two sub-questions need responses.

5.2.3.1 What impact does the time and number of repetitions of the 45 target words, 1 to 8 times for the RMCG and 10 times for the GRG, have on word learning and retention?

First, the FWMs of the RMCG partial correlation with the LDT AS and GFT scores were analysed, followed by an analysis of the BWMs of the RMCG partial correlation with the LDT AS and GFT scores. Then, CQs of the GRG partial correlation with the LDT AS and GFT scores were analysed.

1.*The FWMs of the RMCG correlation with the LDT and GFT*

The mean score of FWMs per person for all words was computed to assess the relationship between the FWMs made by the RMCG and the LDT AS, and the relationship between the FWMs made by the RMCG and the GFT scores. A partial correlation was computed to assess the relationship between the FWMs for list A in the first week and the first LDT AS, and between the FWMs for list A in the first week and the first GFT scores. When controlling the VST scores, no significant relationship was found between the FWMs and the first LDT AS. However, there was a significant relationship between the FWMs and the first GFT scores (see Table 5.42).

Table 5.42 *Partial correlation results of the relationship of FWMs of the RMCG and the first LDT AS and the FWMs of the RMCG and the first GFT scores for list A*

	r	***df***	***p***
LDT 1 AS	.280	22	.185
GFT 1 scores	.440	22	.03

The second partial correlation was computed to assess the relationship between the FWMs for lists A and B in the second week and the second LDT AS and GFT scores. Furthermore, a third partial correlation was computed to assess the relationship between the FWMs of lists A, B, and C in the third week and the third LDT AS and GFT scores. Also, a fourth partial correlation was computed to assess the relationship between the FWMs of list A, B, and C in the third week and the fourth LDT AS and the fourth GFT scores. When controlling the VST scores, these correlation analyses failed to reach significance.

2. The correlation between BWMs and LDT AS and GFT scores

The mean score of BWMs per person for all words was computed to assess the relationship between the BWMs of the RMCG and the LDT AS, and the relationship between the BWMs of the RMCG and the GFT scores. A partial correlation was computed to assess the relationship between the BWMs for list A in the first week and the first LDT AS and GFT scores. Following this a second partial correlation was computed to assess the relationship between the BWMs of lists A and B in the second week and the second LDT AS and GFT scores. Then, a third partial correlation was computed to assess the relationship between the BWMs of lists A, B, and C in the third week and the third LDT AS and GFT scores. Then, a fourth partial correlation was computed to assess the relationship between the BWMs of lists A, B, and C in the third week and the fourth LDT AS and GFT scores. When controlling the VST scores, these correlation analyses failed to reach significance.

3. The correlation between CQs scores of the GRG and the LDT AS and the GFT scores

The mean score of the CQs per person for all words was computed to assess the relationship between the CQs score for the GRG and the LDT AS, and the relationship between the CQs score for the GRG and the GFT scores. A partial correlation was computed to assess the relationship between the CQs score for list A in the first week and the first LDT AS and GFT scores. Then, a second partial correlation was computed to assess the relationship between the CQs score for list A in week 1, plus the CQs score for list B in week 2 and the second LDT AS and GFT scores. Also, a third partial correlation was computed to assess the relationship between the CQs score for list A in week 1, plus the CQs score for list B in week 2, plus CQs score for list C in week 3 and the CQs score and the third LDT AS and GFT scores. Finally, a fourth partial correlation was computed to assess the relationship between the CQs score for list A in week 1, plus the CQs score for list B in week 2, plus the CQs score for list C in week 3 and the CQs score and the fourth LDT AS and GFT scores. When controlling the VST scores, these correlations analyses failed to reach significance.

5.2.3.2 Do time of learning and spacing affect learning from word cards and graded readers?

1. ***The impact of learning time on the RMCG performance***

According to Leitner (1972), in the *Leitner system* the learner follows the process of moving flashcards forward and when a card reaches the fifth section this indicates that the word card is learned and the learner did not need to repeat this card (as cited in Godwin-Jones, 2010). Regarding FWMs and how they were recorded, as elaborated on in the Methodology Chapter, every time a participant moved a card to the fifth section it was considered to be learned, and therefore this was recorded as the time of learning for each target word. The percentage of the number of participants learning each target word each day of learning was calculated. This process of recording was applied for each day in each week of learning, and for each target word in the three lists.

In this section, the target words recorded as learned by high numbers of participants in the early days of learning were considered to be easily learned. However, the target words

learned by the end of the week, and which might remained in the *Leitner system* when another list was introduced, were considered difficult to learn. For the easily learned words, and for those difficultly learned words, an association with the average of recognition in the LDT and the average of production in the GFT at the end of the fifth day of each week is addressed.

a. Easily learned words in the first week

In the first week of learning the words *ingenious, new-fangled, shrewd, trespass,* and *banish* were learned by the end of the fifth sessions by 100% of the RMCG participants. The word *shrewd,* in particular, was learned by all the participants by the fourth session and more than half the group (i.e., 52%) managed to learn it by the second session. Although most participants learned these words in session 3 and session 4, these target words showed the highest numbers of participants' recognition in the LDT and the GFT as can be seen in Appendix 6 (N). The number of participants who recognised the target words *ingenious, new-fangled, shrewd, trespass,* and *banish* were 92%, 100%, 96%, 92%, and 92% respectively. Moreover, the number of participants who produced the target word *ingenious, new-fangled, shrewd, trespass,* and *banish* were 88%, 96%, 92%, 92%, and 100% respectively.

b. Words found difficult to learn in the first week

Some of the participants did not move the words *vicissitude* (i.e., 16%)*, proprietorship* (i.e., 8%), *postulation* (i.e., 20%)*, contentious* (i.e., 12%)*, elucidate* (i.e., 12%)*, detestable* (i.e., 8%) and *pertinent* (i.e., 20%) to the fifth section by the fifth session, so started the second week with target words from list A still in the Leitner box (see Figure 5.4). Despite this, 84% only managed to move the word *vicissitude* to the fifth section by the fifth session, the number of participants who recognised it in the LDT 1 was 88%, and who produced it in the GFT 1 was 96%, which was relatively high. However, the words *proprietorship, postulation, contentious, elucidate, detestable,* and *pertinent* were recognised by a lower number of participants in the LDT AS and the GFT scores, as illustrated in Appendix 6 (Q 1).

When all participants moved the previous target words to the fifth section by the second session in the second week, they performed better in the weekly immediate post-tests. The number of participants who recognised the words *vicissitude, proprietorship, postulation, elucidate, detestable,* and *pertinent* in the second week was higher in LDT 2 than in LDT 1. The number of participants who produced the words*, proprietorship, contentious, elucidate, detestable,* and *pertinent* in the second week was higher in GFT 2 than GFT1. These figures are

illustrated in Appendices 5 (N 1 and Q 1). Despite the fact that the participants finished learning these words on the second day of the week, the number of participant recognition and production was higher in the second weekly immediate post-test (see Appendix 6 (Q 1)).

The percentage of participants who were able to move target words from list A to the fifth section by the end of the first week was computed to assess the relationship between the percentage of participants who finished learning target words from list A and the percentage of participants who recognised each target word in the LDT on the one hand, and with the percentage of participants who produced each target word in the GFT on the other hand.

There is a significant positive correlation between the percentage of participants who managed to learn target words from list A in the first week and the percentage of participant recognition of the target word in the first LDT (r = .768, n =15, p = 001). There was also a significant positive correlation between the percentage of participants who managed to learn target words from list A in the first week and the number of participant production of the target words in the first (GFT, r = .498, n =15, p = .05).

c. Easily learned words in the second week

In the second week of learning, the words *stupendous, gargantuan, acrid, obstinate, colossal, bonny,* and *unequivocally* were learned by the end of the fifth sessions by 100% of the RMCG participants (see Figure 5.6). The word *bonny,* in particular, was learned by 96% of the participants by the fourth session, and 48% of the group managed to learn it by the third session. Although most participants learned these words in session 3 and session 4, these target words show the highest number of participant recognition in the LDT and in the GFT as well. The number of participants who recognised the words *stupendous, gargantuan, acrid, obstinate, colossal,* and *bonny* was 80%, 88%, 84%, 88%, 92%, and 84% respectively (see Appendix 6 (O 1). Moreover, the number of participants who produced the words *stupendous, acrid, obstinate,* and *bonny* was 72%, 72%, 76%, and 84% respectively. The words *gargantuan* and *colossal,* however, were produced by a lower number of participants (i.e., 56% and 68%, respectively) (see Appendix 6 (R 1)).

d. Words found difficult to learn in the second week

Although the word *unequivocally* was considered to be learned by 100% of the participants by the end of the week, the numbers of participant recognition (i.e., 76%) and production (i.e., 52%) were relatively low. The reason might be that 92% of the participants

moved the word *unequivocally* by the fourth session and might had forgotten it when they received the weekly immediate post-test by the fifth session.

Some of the participants did not move the words *adroit* (i.e., 8%), *misgiving* (i.e., 4%), *irksome* (i.e., 4%), *mystification* (i.e., 10 %), *ebullient* (i.e., 4%), *idiosyncratic* (i.e., 20%), *apprehensive* (i.e., 4%), or *brittle* (i.e., 4%) to the fifth section by the fifth session, so they started the third week with these words in the Leitner box (see Figure 5.6 and Figure 5.8). Despite the fact that 80% only managed to move the word *idiosyncratic* to the fifth section by the fifth session, the percentage of participants who recognised it in the LDT 2 was 92% (see Appendix 6 (O 1)). However, the number of participants who produced it in GFT 2 was relatively low (i.e., 48%). The lowest percentages of participant recognition in LDT 2 for the words *misgiving* and *mystification* were 64% and 76%, respectively. As for the percentage of participant production in GFT 2 for the word *mystification,* this was relatively low (i.e., 44%), and for the word *misgiving,* the percentage of participant production in GFT 2 was instead relatively high (i.e., 80%) (see Appendix 6 (R 1)).

When 100% of the participants moved the words *adroit*, *misgiving*, *irksome*, *mystification*, *ebullient*, *idiosyncratic*, *apprehensive,* and *brittle* to the fifth section by the second session in the third week they performed better in the weekly immediate post-tests for some of the target words. The number of participants who recognised the target words *adroit, misgiving,* and *mystification* in the third week was higher in LDT 3 than in LDT 2. The number of participants who produced the target words *adroit, irksome, mystification, ebullient,* and *idiosyncratic* in the third week was higher in GFT 3 than GFT 2. These figures are illustrated in Appendix 6 (O 1 and R 1). Despite the fact that the participants finished learning these words on the second day of the third week, the numbers of participant recognition and production were higher in the third weekly immediate post-test (see Appendix 6 (O 1 and R 1)).

The percentage of participants who were able to move target words from list B to the fifth section by the end of the second week was computed to assess the relationship between the percentage of participants who finished learning target words from list B and the percentage of participants who recognised each target word in the LDT on the one hand, and with the percentage of participants who produced each target word in the GFT on the other hand.

There was no relationship between the percentage of participants who manage to learn target words from list B in the second week and the numbers of participant recognition of the target words in the second week LDT (r = .407, n =15, p =.132). Moreover, there was no significant relationship between the percentage of participants who manage to learn target words from list B in the second week and the numbers of participant production of the target words in the second GFT (r = .284, n =15, p = .305).

e. Easily learned words in the third week

In the third week of learning, the words *tedious, astound, veritably, feeble, sluggishly, lucid,* and *folly* were learned by the end of the fifth sessions by 100% of the RMCG participants. The words *feeble* and *folly,* in particular, were learned by 96% of the participants by the fourth session (see Figure 5.8). Although most participants learned these words in session 3 and session 4, some of these target words showed the highest number of participant recognition in the LDT and the GFT as well (see Appendix 6 (P 1 and S 1). The percentage of participants who recognised the words *astound, sluggishly, lucid,* and *folly* was 76%, 72%, 84%, and 88%, respectively. Moreover, the percentage of participants who produced the target words *tedious, astound, veritably, sluggishly, lucid,* and *folly* was 60%, 76%, 68%, 64%, 64%, and 68%, respectively. Although the word *feeble* was considered to be learned by 100% of the participants by the end of the week, the percentage of participant recognition (i.e., 64%) and production (i.e., 56%) was relatively lower than for other target words.

f. Words found difficult to learn in the third week

Some of the participants did not move the words *grievous* (i.e., 12%), *cogent* (i.e., 12%), *toil* (i.e., 4%), *expeditious* (i.e., 4%), *afflicted* (i.e., 4%), *petrify* (i.e., 4%), *perseverance* (i.e., 16%), or *rambunctious* (i.e., 4%) to the fifth section by the fifth session (see Figure 5.8). Despite the fact that 84% only managed to move the word *perseverance* to the fifth section by the fifth session, the number of participants who recognised it in LDT 3 was 84%. However, the number of participants who produced it in GFT 3 was relatively low (i.e., 44%). The lowest percentage of participant recognition in LDT 3 was for the words *grievous* and *cogent,* where 52% and 56%, respectively, of the participants produced them in GFT 3 (see Appendix 6 (P1 and S1)).

The percentage of participants who were able to move target words from list C to the fifth section by the end of the third week was computed to assess the relationship between the percentage of participants who finished learning target words from list C and the percentage of participants who recognised each target word in the LDT on the one hand, and with the percentage of participants who produced each target word in the GFT on the other hand.

There was no significant relationship between the percentage of participants who managed to learn target words from list C in the third week and the numbers of participant recognition of the target words in the third LDT ($r = .00$, $n = 15$, $p = 1.00$). However, there was a positive and significant correlation between the percentage of participants who managed to

learn target words from list C in the third week and the number of participants producing the target words from list C in the third GFT (r = .510 n =15, p = .05).

The percentage of participants who moved the word cards to the fifth section in the *Leitner system* in the three weeks was computed to compare the percentage of participants finishing the task for each target word with the percentage of participants' recognition in LDT 4 and the percentage of participant production in GFT 4 as well. 100% of the participants successfully moved the target words in lists A and B, and most of the target words in list C, to the fifth section in the *Leitner system*. Although list A was learned earlier in the first week and in the first two days of the second week, as illustrated in the previous sections, the numbers of participant recognition in LDT 4 and production in GFT 4, which was about four weeks later, were higher than the numbers of participant recognition and production of list B and C. Also, the numbers of participant recognition in LDT 4 and production in GFT 4 were higher for list B than list C. (See Appendices 6 (N1, O1, and P1) for the RMCG recognition of list A, B, and C) and (see Appendices 6 (Q 1, R 1, S1) for the RMCG production of list A, B, and C).

The percentage of participants who were able to move target words from list, A, B, and C to the fifth section by the end of the third week was computed to assess the relationship between the percentage of participants who finished learning target words from lists A, B, and C and the percentage of participants who recognised each target word in LDT 4 on the one hand, and with the percentage of participants who produced each target word in GFT 4 on the other hand.

There was a positive and significant correlation between the percentage of participants who managed to learn target words from lists A, B, and C in the three weeks and the numbers of participant recognition of the target words in LDT 4 (r = .420, n = 45, p = .004). There was a positive and significant correlation between the percentage of participants who managed to learn target words from lists A, B, and C in the three weeks and the numbers of participant production of the target words from lists A, B, and C in GFT 4 (r = .354 n =45, p = .017).

2. The impact of space of between occurrence of target word in the graded readers on the GRG performance

As explained in the Methodology Chapter, each of the graded readers were divided into five texts, and each text consisted of 5% target words. Table 5.43 shows the spacing between each target word and the mean number of participant recognition in the immediate LDT and post-test LDT, and the mean of participant production in the immediate GFT and the post-test GFT. Target words with all types of spacing had very close numbers of participant recognition

in the immediate LDT and post-test LDT, and participant production in the immediate GFT and the post-test GFT. Therefore, spacing in the GRG did not affect target word recognition and production.

Table 5.43 Spacing of target words in the graded readers and mean of percentages of participant recognition in the immediate LDT and the post-test LDT, and the mean of percentages of participant production in the immediate GFT and the post-test GFT

spacing	**No. of target words (Max = 45)**	**immediate LDT**	**immediate GFT**	**post-test LDT**	**post-test GFT**
consecutive repetition	14	74%	56%	68%	32%
one day spacing	17	76%	59%	70%	35%
two days spacing	8	73%	59%	69%	37%
three days spacing	6	71%	54%	67%	31%

As for timing of the occurrence of target words in the graded readers, it is noticeable that target words occurring in the second and third session showed a higher percentage of participant recognition in the immediate LDT and in the post-test LDT, and a higher percentage of participant production in the immediate GFT and in the post-test GFT. Target words that occurred in the first and second graded readers' texts had the lowest percentage of participant recognition in the immediate LDT and the post-test LDT, and the lowest percentage of participant production in the post-test GFT (see Table 5.44).

Table 5.44 *Timing of target words in the graded reader and means of percentages of participant recognition in the immediate LDT and post-test LDT and the means of percentages of participant production in the immediate GFT and the post-test GFT*

timing in text	No. of target words (Max = 45)	immediate LDT	immediate GFT	post-test LDT	post-test GFT
1st and 2nd	2	54%	54%	58%	24%
1st and 3rd	5	76%	61%	76%	35%
1st and 4th	6	71%	66%	67%	41%
1st and 5th	6	71%	54%	67%	31%
2nd and 3rd	7	82%	63%	73%	37%
2nd and 4th	7	77%	57%	67%	34%
2nd and 5th	2	78%	38%	72%	22%
3rd and 4th	2	57%	42%	70%	30%
3rd and 5th	5	75%	60%	67%	37%
4th and 5th	3	80%	51%	60%	27%

5.2.4 The relative contribution of different factors to the learning and retention of target words

All and all, different predictors which might contributed to the learning of the target words across the three groups were analysed and tested in a number of correlation and regression analyses. A number of predictors were shared by all groups, whilst others were only applicable to one group, but not another. Predictors shared by all groups were: difficult sounds for Arab ESL learners in a target word; the number of syllables in a target word; the number of letters in a target word; complexity and derivational morpheme frequency in a word; and the number of difficult consonant and vowels for Arab ESL learners in each target word. In addition to these predictors, for the RMCG their frequency of exposure to a target word and the FWMs and the BWMs of word cards were potential predictors, whilst for the GRG their frequency of exposure to a target word and the spacing of the target words in the graded readers were potential predictors. Finally, for the CG, their frequency of exposure to a target word was a potential predictor.

The frequency of exposure was measured as follows. Each of the three groups were exposed to target words at different frequencies. The RMCG was exposed to list A five times in the first week and allowed to have list A in Leitner boxes through the second and third weeks,

therefore, the RMCG was exposed to list A 15 times (i.e., 5+5+5 =15 times) and, when adding the occurrence of target words from list A in the LDT and GFT, they were exposed to list A eight times in the LDTs and GFTs (i.e. 2+2+2+2=8 times), so the total was 23 times for the RMCG for list A. When conducting similar calculations for lists B and C, the RMCG were exposed to words from list B 16 times ((5+5) + (2+2+2) =16), and to words from list C nine times ((5) +(2+2) = 9). These figures represented the possible exposure as many participants finished learning target words earlier. Therefore, the frequency of exposure for the RMCG ranged from nine to 23 times.

As for the GRG, the frequency for the three lists also differed. The GRG was exposed to list A ten times in the first week (i.e., five times in each text of *the Free Culture* twice a week) and they were exposed to list A eight times in the LDTs and GFTs (i.e., 2+2+2+2 = 8), so the total exposure was 18 times. When calculating the exposure to lists B and C the total exposure to the target words for list B was 16 times (i.e., (5x2 in the second week) + (2+2+2 = 6)), and for list C the total exposure was 14 times (i.e., (5x2 in the third week) + (2+2 = 4)). Therefore, the frequency of exposure for the GRG ranged from 14 to 18.

The CG exposure to list A was eight times in the LDTs and GFTs (i.e., 2+2+2+2 = 8 times), to list B was six times (i.e. 2+2+2 = 6 times), and to list C was four times (i.e., 2+2 = 4 times). Therefore, the frequency of the CG exposure to target words ranged from 4 to 8.

In addition to frequency of exposure, the number of syllables for each of the target words was considered to be one of the predictors (see Appendix 6 (H)). The number of letters for each target word were counted and target words classified according to the number of letters. As for complexity predictors, target words were classified according to their category of complexity: simple; complex; compound; or compound and complex. Target words were also categorised into three groups: low frequency; mid frequency; and high frequency, as illustrated in Appendix 6 (G 1). Means of FWMs and BWMs, seen in Appendix 6 (C 1 and C 2), were used to classify target words. Target word were spaced in the graded readers in a different way, as illustrated in Appendix 5 (U 1). As for difficult sounds for Arab ESL learners, sounds that were difficult for Arab ESL learners, such as /p/, /v/, and /g/, (see Appendix 6 (I)) and short vowels, such as /ʌ/, /ɛ/, /a/, /ɪ/, /ɒ/, /i/, and /ʊ/, (see Appendix 6 (J)) were counted in each target word. Table 5.45 shows the predictors of this multiple regression analysis followed by a brief explanation of each of the predictors and the results of the regression and correlation analysis.

Table 5.45 *Predictors of this multiple regression analysis and their measurements*

predictors	**frequency of exposure to a target word**	**number of syllables in a target word**	**number of letters of each target word**	**complexity**	**derivational morpheme (dm) frequency of a target word**
scale	Number of exposure from four to 23 times	number of syllables from one to five	Number of letters from five to 18	simple target word = 1 complex target words = 2 compound target words = 3 compound and complex target words = 4	words with no frequency dm = 0. words with high frequency dm = 1 words with mid frequency dm = 2 words with low frequency dm = 3
Predictors	**FWM**	**BWM**	**spacing of the target words in the graded readers**	**difficult sounds for Arab ESL learners**	
scale	means of FWMs	means of BWMs	from zero to three days	number of difficult consonants and vowels from zero to six.	

First of all, correlations were computed between the variables mentioned in Table 5.45 and the outcome variables for the current study. The outcome variables were the LDT 4 AS and GFT 4 scores. Only significant correlations are reported. Subsequently, a regression analysis was carried out with the predictors that correlate significantly with the dependent variables. In the first instance, analyses were carried out with either the total LDT AS and the total GFT scores for all groups as the dependent variable, or with the LDT AS and the GFT scores for each of the groups separately.

1. *The analysis of the three groups*

Correlations were computed among seven self-concept scales on the data for 45 target words. The results suggested that there was a negative and significant correlation between the

GFT scores for all groups and the number of syllables in the target word for the GRG, (r = -.348, n = 45, p = .019). There was also a negative and significant correlation between the GFT scores for all groups and the number of letters in the target word, (r = -.329, n = 45, p = .027) (see Table 5.46).

Multiple regression analysis was used to test if the number of syllables in the target word and the number of letters in the target word predicted the GFT scores for all groups. The results of the regression indicate that one predictor, namely the number of syllables, explained 12 % of the variance (R^2=.12, F (1,43) = 5,90, p =.019). The analysis showed that the number of syllables in the target word significantly predicted the GFT scores for all groups (β = -.34, t (44) = -2.430, p = 019) (see Figures 1 and 2 in Appendix 6 (V)).

Table 5.46 *Correlations between word characteristics and LDT AS and GFT scores for all groups*

	LDT AS	GFT scores	number of syllables	complexity	number of letters	difficult consonants and vowels	derivational morpheme frequency
LDT AS	1	.586** p < .0001	-.050 p = .744	-.057 p = .706	-.116 p = .448	-.016 p = .917	-.161 p = .292
GFT scores		1	- .348* p = .019	-.153 p = 318	-.329* p = .027	-.234 p = .101	-.196 p = .197
number of syllables			1	.564** p < .0001	.897** p < .0001	.696** p < .0001	.503** p < .0001
complexity				1	.674** p < .0001	.524** p < .0001	.581** p < .0001
number of letters					1	.626** p < .0001	.683** p < .0001
difficult consonant and vowels						1	.316* p = .034

**. Correlation at 0.01(2-tailed). *. Correlation at 0.05(2-tailed).

Correlations were computed among LDT AS and GFT scores for all groups and frequency of exposure (see Table 5.47).

Table 5.47 *Correlations between frequency of exposure and LDT AS and GFT scores for each group separately*

	LDT AS	GFT scores	RMCG frequency of exposure	GRG frequency of exposure	CG frequency of exposure
LDT AS	1	586** $p < .0001$	.312* $p = .037$	312* $p = .037$	312* $p = .037$
GFT scores		1	.348* $p = .019$	.348* $p = .019$	.348* $p = .019$
RMCG frequency of exposure			1	1.00** $p < .0001$	1.00** $p < .0001$
GRG frequency of exposure				1	1.00** $p < .0001$

**. Correlation at 0.01(2-tailed).*. Correlation at 0.05(2-tailed).

2. The analysis of each group

There was a positive and significant correlation between the LDT 4 AS of the RMCG and the frequency of exposure to a target word for the RMCG ($r = .625$, $n = 45$, $p < .0001$). Simple linear regression analysis was used to test if frequency of exposure to a target word predicted the LDT 4 AS for the RMCG. The results of the regression indicate that this predictor explained 39% of the variance (R^2=.39, F (1, 43) = 27.51, $p < .0001$). The analysis showed that the frequency of exposure to the target word significantly predicted LDT 4 AS of the RMCG

*(*β = .62*, t* (44) = 5.245, *p* <.0001)[2]. (see Figures 3 and 4 in Appendix 6 (V)). This means, for the RMCG, the higher the frequency of exposure to the target word, the easier the recognition of the target word was in the LDT.

There was a positive and significant correlation between the GFT 4 scores of the RMCG and the frequency of exposure to a target word for the RMCG (*r* = .476, *n* = 45*, p* = .001). Simple linear regression analysis was used to test if frequency of exposure to a target word predicted the GFT 4 scores for the RMCG. The results of the regression indicate that one predictor explained 22% of the variance (R^2=.22, *F* (1,43) = 12,61, *p* =.001). It was found that frequency of exposure to a target word for the RMCG significantly predicted GFT 4 scores for the RMCG (β = .47, *p* <.0001). The analysis shows that the number of syllables significantly predicted GFT 4 scores for the GRG *(*β = -.46*, t* (44) = -3.405, *p* = .001) (see Figures 5 and 6 in Appendix 6 (V)). This means, the more syllables in a word, the harder it is to produce in the GFT.

As for the GRG, there was a negative and significant correlation between the GFT 4 scores and the number of syllables in target words (*r* = -.461, *n* = 45*, p* = .001). Also, there was a negative and significant correlation between the GFT scores for the GRG and the number of letters in a target word (*r* = -.403, *n* = 45*, p* = .006). This means that longer words (as measured in either syllables or letters) were retained less well than shorter words. Moreover, there was a negative and significant correlation between the GFT 4 scores and the number of difficult consonants and vowels in the target word (*r* = -.352, *n* = 45*, p* = .018). Simple linear regression analysis was used to test if the number of syllables in the target word, the number of letters in the target word, and difficult consonants and vowels for Arab ESL learners in the target word predicted the GFT scores for the GRG. The results of the regression indicate that only the number of syllables explains 21% of the variance (R^2=.21, *F* (1,43) = 11,59, *p* =.001). It was found that the number of syllables in the target words predicted the GFT scores of the GRG (β = -.46, *p* =.001). The analysis shows that the number of syllables significantly predicted GFT scores of the GRG *(*β = -.46*, t* (44) = -3.405, *p* = .001) (see Figures 7 and 8 in Appendix 6 (V)).

---

[2] The histogram of standardized residuals indicates that random errors are approximately normally distributed, as does the normal P-P plot of standardized residuals, which shows points that are not completely on the line, but close.

5.3 Summary of Findings

The results suggested that intentional learning using word cards and the *Leitner system* led to better recognition and production of low frequency target words than did incidental learning from graded readers in both the short and long term. The RMCG out-performed the GRG and the CG in each of the weekly immediate post- tests and in the delayed post-test. The results show a significant difference in the RMCG performance in recognising target words in the LDT and in producing target words in the GFT in each of the weekly immediate post-tests and in the delayed post-test. The GRG did, however, outperformed the CG, and showed able to recognise and produce a considerable number of the target words in the LDT and in the GFT in each of the weekly immediate post-tests and the delayed post-test.

Both the RMCG and the GRG were able to recognise and produce more than half of the target words at the end of the learning phase, although there was a gradual, but steady decline for the RCMG and the GRG performance from test one to test three in the weekly immediate post-tests, and from the weekly immediate post-tests to the delayed post-test.

In general, there was a gradual but steady decrease in RTs of the RCMG and of the GRG from test one to test three in the weekly immediate post-tests, and from the weekly immediate post-tests to the delayed post-test.

The results also suggested that the word cards and the *Leitner system* were an effective technique for intentional vocabulary learning, and the graded readers were an effective technique for incidental vocabulary learning. It was possible to follow the progress of the intermediate Arab ESL learners of the 45 target words not only after learning, but while using word cards and the *Leitner system* or graded readers. Moving word cards forward to the fifth section for the RMCG, and answering the CQs correctly by the end of each graded readers' text for the GRG, each provided evidence of learning while using the techniques. The word card movement analysis for the RMCG and the CQs responses analysis for the GRG showed that the RMCG was learning from word cards and the GRG was noticing target words from the graded readers.

The RMCG retained most of the 45 target words and the analysis of retention in the first, second, and third weeks of learning show successful retention of each of the lists. The RMCG retained all target words in lists A and B by the subsequent week of learning by repeating all words five times or more. However, the RMCG managed to move most word cards from list C five times or more in the third week of learning. Some target words out of the fifteen

in list C were not moved to the fifth section by 100% of the participants, but the RMCG successfully retained most of list C target words.

Conversely, the GRG was introduced to list A in week 1, to list B in week 2, and to list C in week 3. The GRG was not able to mix the words from all the three lists due to the design of the graded readers. The GRG was presented with each word twice on two different days and they successfully gave the correct responses to about 73% of the CQs in list A, list B, and list C, respectively.

The L2 learners' performances decreased when large quantities of vocabulary input were maintained over an extended and consistent period. The RMCG recognition of target words in the LDT steadily declined from the first immediate post-test to the delayed post-test. However, the GRG target word recognition declined gradually from the first weekly immediate post-test to the third weekly immediate post-test, but that recognition increased in the delayed post-test. The CG recognition performance was different from the RMCG and the GRG because their recognition rate steadily increased from the first weekly immediate post-test to the delayed post-test.

As for the GFT, the production of both groups declined from the first weekly immediate post-tests to the delayed post-test. The RMCG production performance steadily declined from the first weekly immediate post-test to the delayed post-test. The GRG production performance steadily declined from the first weekly immediate post-test to the third immediate post-test, but there was a sharp drop in production of target words from the third weekly post-test to the delayed post-test. The CG failed to produce target words in the GFT and their scores were, therefore, extremely low in comparison with the RMCG and the GRG.

There was a significant difference between the performance of the RMCG in comparison to the GRG in the weekly immediate post-tests (except for week 1). Furthermore, both groups differed significantly from the CG in the weekly immediate post-tests.

An important finding is that there was a positive and significant correlation between the percentage of the RMCG participants who moved target words from lists A, B, and C forward to the fifth section in the Leitner box and the percentage of the RMCG participant recognition of lists A, B, and C in LDT 4. Moreover, there was a positive and significant correlation between the percentage of the RMCG participants who moved target words from lists A, B, and C forward to the fifth section of the Leitner box and the percentage of the RMCG participant production of lists A, B, and C in GFT 4. Conversely, there was no clear indication that spacing and timing in the GRG played an important role in the recognition of target words in the LDT and their production in the GFT.

In a qualitative analysis, the effect of target word length and frequency of morphological derivational morphemes were analysed. The analysis shows higher numbers than the average

for recognition by the RMCG, and that the GRG recognised and produced a considerable number of long and short target words. It also shows that the RMCG and the GRG had difficulty in recognising and producing some short target words and long target words. The RMCG and the GRG highly recognised and produced many target words with high, mid, and low frequency derivational morphemes. The RMCG and GRG also had difficulty in recognising target words with high, mid, and low frequency derivational morphemes.

Furthermore, in the qualitative analysis addressed the effect of special difficulty of Arab ESL learners in target word the RMCG and the GRG highly recognised target word with difficult consonants, short vowels and consonant cluster. These characteristics might cause difficulty in recognising or producing some of the target words. The RMCG and GRG reported difficulty learning some target words with special phonological characteristics.

Additionally, the multiple regression analysis shows that the frequency of exposure to target words predicted the LDT AS and GFT scores for the RMCG. Also, the multiple regression analysis shows that the number of syllables in a target word predicted the GFT scores for the GRG and for all the groups combined.

Finally, the results provided evidence of the superiority of intentional learning using word cards in conjunction with the *Leitner system* over incidental learning using graded readers. The overall performance of the RMCG was superior to the GRG during the learning process, in the weekly immediate post-tests, and in the delayed post-test. The GRG was also superior to the CG and obtained better results in the LDT and the GFT. Even though there was a decrease in the learners' performance for both the RMCG and the GRG, both groups retained a

considerable number of target words in the delayed post-test after two weeks from learning. See Table 5.48 for overall comparison between groups.

Table 5.48 *Summary of main findings to compare between groups*

	RMCG	**GRG**	**CG**
week 1 recognition out of 15	13.05 (SD = 6.75)	12.45 (SD = 7.96)	5.10 (SD = 18.00)
week 2 recognition out of 30	25.50 (SD = 7.00)	21.60 (SD = 9.66)	12.30 (SD = 11.20)
week 3 recognition out of 45	35.55 (SD = 7.05)	30.60 (SD = 9.96)	17.10 (SD = 5.68)
delayed post-test recognition out of 45	33.75 (SD = 7.38)	31.50 (SD = 6.38)	20.25 (SD = 6.93)
week 1 production out of 15	11.70 (SD = 14.01)	8.85 (SD = 10.55)	-
week 2 production out of 30	22.50 (SD = 10.48)	17.10 (SD = 11.52)	-
week 3 production out of 45	30.60 (SD = 9.04)	18.90 (SD = 7.74)	-
delayed post-test production out of 45	25.65 (SD = 8.88)	15.30 (SD = 6.67)	-
week 1 LDT RT (*ms*)	2354.97 (SD = 256.82)	2756.33 (SD = 265.27)	2824.10 (SD = 949.73)
week 2 LDT RT (*ms*)	2238.13 (SD = 205.03)	2413.40 (SD = 226.17)	2248.52 (SD = 171.98)
week 3 LDT RT (*ms*)	2118.10 (SD = 263.49)	2277.90 (SD = 236.37)	1758.41 (SD = 230.50)
delayed post-test LDT RT (*ms*)	2073.00 (SD = 152.95)	1785.60 (SD = 154.40)	1700.44 (SD = 205.81)
mean of FWMs	5. 19 (SD = .12)	-	-
mean of BWMs	.21 (SD = .13)	-	-
mean of CQs score (Max = 90)	-	68.80 (SD = 4.49)	-

6 Discussion

6.1 Overview

The study set out to assess the impact of intentional learning using word cards in conjunction with the *Leitner system* and incidental learning using graded readers on the learning of 45 target words over a period of three weeks. It investigated the performance of two groups of Arab ESL learners who were learning target words by employing the spacing repetition technique. The performances of the groups (i.e., the RMCG and the GRG) were compared with each other and with a control group who did not receive any vocabulary learning guidance. The accumulative learning of target words was tested in the three weekly immediate post-tests and in the delayed post-test after two weeks from learning. The approximate number of repetitions and the degree of spacing necessary for intentional learning, using word cards in conjunction with the *Leitner system,* and incidental learning, using the graded readers, were investigated in both the RMCG and the GRG. Difficult word characteristics were analysed in relation to the three groups' recognition and production of target words. Furthermore, the learning burden effect and the effect of decay on retention were investigated. The following provides a summary of the results, followed by an interpretation of the findings in relation to the research questions posed and the previous research.

6.2 A Summary of the Results

Overall the RMCG outperformed the GRG and the CG in recognizing and producing the target words. The results of the weekly immediate post-tests and the delayed post-test clearly show that the intentional learning, using word cards in conjunction with the *Leitner system,* was more beneficial than the incidental learning, using graded readers, for immediate retention and in long-term learning. The LDT AS and the GFT scores also clearly show that graded readers helped Arab ESL learners in the intermediate language proficiency level to learn target words incidentally. The GRG outperformed the CG in target word recognition and production. It is important to highlight that the RMCG and the GRG were successful because

they were able to recognize and produce more than half of the target words after the learning phase had finished.

Target word recognition was found to be faster for the RMCG than for the GRG and the CG. The successful recognition rate, as measured by short RTs, indicates that participants retained the form of the target words. The RT of the RMCG and the GRG gradually declined from the first weekly immediate post-test to the delayed post-test. The participants, namely the RMCG and the CG, showed recognition of the words through shorter RTs from the first weekly immediate post-test to the delayed post-test (see Figure 5.1).

The result of the correlation analysis of FWMs and BWMs made by the RMCG and their post-test revealed that learning took place from word cards. The significant correlation between the number of participants who managed to move target words from lists A, B, and C to the fifth section in the *Leitner system* and the numbers of participant recognition of the target words in the delayed post-test indicates the efficiency of word cards. Furthermore, the positive and significant correlation between the number of participants who moved target words from lists A, B, and C to the fifth section of the *Leitner system* and the numbers of participant production of the target words from lists A, B, and C in the delayed post-test indicates the efficiency of word cards.

The qualitative analysis of the effect of word characteristics shows how target words with common difficult characteristics and those with special difficult characteristics for Arab ESL learners were recognised and produced. The RMCG managed to recognise and produce target words, despite the difficulty of many of them. Many target words that were recognised and produced better than other target words were long words, words with low frequency derivational morphemes, words with difficult consonant sounds, and words with short vowels. This shows the extent to which spaced learning using word cards and the *Leitner system* and graded readers were effective.

Finally, the results suggest that spaced learning via intentional learning using word cards and the *Leitner system,* or incidental learning using graded readers helped Arab ESL learners learn and retain most of the 45 target words. The intermediate Arab ESL learners in Albaha University retained target words using word cards with the *Leitner system* or graded

readers and could give evidence of learning. The remainder of the chapter is structured according to the research questions given in the Literature Chapter.

6.3 Discussion Relating to Each Research Question

6.3.1 Which mode of learning results in superior retention during and after fifteen days of learning?

The results of the study seem to constitute support for Hypothesis I which suggests that the use of word cards results in superior retention by comparison with graded readers. This study responded to the debate about whether intentional or incidental learning is more effective. This debate has been addressed by many researchers, such as Hulstijn (2003), Nation, (2011) and Schmitt (2000) as they discuss the dispute related to whether learners intentionally retain L2 vocabulary or rely on retaining L2 vocabulary incidentally through reading. The results provided evidence of the benefit of intentionally learning L2 vocabulary through the use of words cards in conjunction with the *Leitner system,* as well as the benefit of graded readers as a means of incidental vocabulary learning. In response to the first research question, the RMCG outperformed the GRG in the number of words retained (see section 5.2.1.5), yet the GRG was able to learn a greater proportion of target words in comparison to previous research (e.g., Waring and Takaki's (2003) study in which learners recognised the form of 6% of 25 new words when repeated from 15 to 18 times after three months from learning).

Thus, the study endorses the findings of researchers who claim that intentional learning results in higher retention rates than incidental L2 vocabulary learning (e.g., Elgort, 2011; Hulstijn, 2003; Nakata, 2008; Nation, 2011; Schmitt, 2000). It had been shown, in both the Pilot study and the Main study, that intentional learning using RM word cards results in better recognition and production of target words in comparison to incidental learning using graded readers.

6.3.1.1 Which of the two groups (the RMCG or the GRG) learn and retain more words as shown by the weekly immediate post-tests and the delayed post-test?

The Hypothesis to this sub-question is that the RMCG will retain more of the 45 target words using word cards than the GRG does using graded readers. The RMCG had the highest LDT AS and the highest GFT scores, the GRG came next, and the CG came last. This sequence of the three groups' LDT AS and GFT scores remained consistent in each of the weekly immediate post-tests and in the delayed post-test. The difference between the first LDT AS for the RMCG and the GRG, and the difference between the LDT AS of the RMCG and the CG and the GRG and the CG were significant in the, first, second and third weekly immediate post-tests and in the delayed post-test. The effect size is low in two comparisons: in the difference between the LDT 1 AS and in the LDT 4 AS between the RMCG and the GRG. However, the effect size between the groups in the four LDT AS comparisons are medium to large.

Performance by the RMCG is expected to be superior to the GRG and the CG as the RMCG received two devices that facilitate vocabulary learning: the word cards and the *Leitner System,* while the GRG had one task to do (i.e. reading a passage). In fact, the GRG had also to respond to comprehension questions after reading and the reading texts were adjusted to the participants' speed of reading in 20 minutes (i.e., 500- to 700-word text length). The RMCG and the GRG had the same amount of time to learn the words and they had two tasks to do, therefore, time-on-tasks was controlled for these two groups (Hill & Laufer, 2006).

6.3.1.2 Performance in word cards in conjunction with the Leitner system

Performance in word cards and the Leitner system

The RMCG learning corresponds to Hulstijn's (2003) characteristics of intentional learning. The participants experienced the process in which they deliberately memorized 45 target words. The LDT AS and the GFT scores for the RMCG indicate that intentional vocabulary learning occurred by focusing on both the form and the meaning conveyed by those words. This contrasts with Krashen (1989) who claims that linguistic knowledge is acquired when learners focus predominantly on the meaning of the items, rather than on the form.

In comparison to Fitzpatrick et al.'s (2008) study, where one subject who learned 400 words intentionally in 20 days was able to recall 54% of them, the RMCG in the current study was exposed to 45 target words in fifteen days (i.e., three weeks, not including the weekends) and were able to recognise 75% and produce 57% of the target words.

Learning the form and meaning of the words confirmed the advantage of word cards over other intentional vocabulary learning techniques, such as memorizing word lists written in one place and using notebooks (Nakata, 2008; Nation, 2013; Waring, 2004). Word cards have many advantages over other vocabulary-learning techniques because of their special *practice retrieval* characteristics (Baddeley, 1997), and because they have the advantages of intentional learning (Mondria & Mondria-De Vries, 1994; Nation, 2001; Schmitt & Schmitt, 1995; Waring, 2004). Learning from word lists might result in a *list effect* (Nakata, 2008), or might result in a focus on learning the first and last words more effectively than the words in the middle (Lewandowsky, 2005). Unlike learning from word lists written in one place, the RMCG classified the words according to their needs, as Nakata (2008) observed in an analysis of the word card movements in the *Leitner system.*

The RMCG experienced being independent in learning vocabulary while using word cards and they did not need the help of an instructor, as suggested by Nation (2001) and Waring (2004). Moreover, the RMCG benefited from the need for conscious concentration while using word cards, as suggested by Mondria and Mondria-De Vries (1994).

The design of the RM cards in this study gave the learners the opportunity to see the target word on one side, then to find the translation, synonym, part of speech, phonetic description, and an example on the reverse side of the card. As suggested by Mondria and Mondria-De Vries (1994), this design of word cards did not allow the RMCG to see the information (e.g., definition/translation) immediately, therefore, the participants tried harder to remember a word and then to check the L2 word information. The participants experienced what Baddeley (1997) calls *retrieval practice* by making the effort of trying to recall a target word when reading one side of the card and allowing for a delay in seeing the word information on the reverse side of the card. This practice resulted in not only successful recall but also in long term retention of vocabulary. Paper-based word cards motivated the RMCG to challenge themselves as they gave them sufficient intervening time to recall the word from memory, as suggested by Nakata (2008).

Adding an example to the reverse side of the cards, adding a translation, and experiencing *retrieval practice* contributed to learning the form of the word and understanding the meaning the target word conveyed in context, as suggested by the results of the GFT. This confirms Waring's (2004) claim that paying attention to the L2 word on one side helps learning receptively, and paying attention to the translation on the other side helps learning the target

word productively. Also, this result lends support to the results of Nakata's study (2008) which found that learning vocabulary with word cards was the most efficient way to learn words. In that study, however, a comparison was made between the use of word lists and word cards (i.e., paper-based and computer-based word cards), and not between word cards and graded readers.

Moreover, the word cards in conjunction with the *Leitner system* allowed the RMCG to repeat target words in the form of overt or silent articulation. This overt or silent articulation of target words, or *subvocal* rehearsal, remained for a short time in the RMCG's WM. The RMCG employed the WM process, as suggested by Dosher (2005), they paid attention to the information on the word cards and were able to manage and alter information according to the demands of the word card and the *Leitner system* task. This process led them to retain the target word in their LTM (Baddeley, 1997).

The RMCG experienced fast learning resulting in long-time retention, as Nation (2013) suggests. He also claims that fast learning does not mean fast forgetting when using word cards. This confirms the speedy learning that Nakata (2008) finds in his study where subjects learned only ten items in ten minutes using word cards. Not only did the RMCG learn target words quickly, but, as Baddeley (1997) suggests, they learned the forms and context for many vocabulary items.

It was demonstrated by the Pilot study that when learners design the cards themselves this might raise their word retention because they will experience success in achieving their goal of memorizing a set of words, as suggested by Mondria and Mondria-De Vries (1994). In the main study, the RM word cards were used for two reasons: firstly, in the pilot the group who learned using RM word cards had slightly higher recognition and production rates than did the group of participants who used FD cards. Secondly, the time the researcher had to ask participants to design cards for 45 target words and to revise them for accuracy and appropriateness of the designs for 25 participants was limited.

The RMCG learned explicitly (i.e., they were consciously learning and they were expected to be tested). They implemented controlled consciousness while learning intentionally from word cards and this is one fundamental characteristic of successful intentional learning, as suggested by many researchers (e.g., Hulstijn, 2003; Nation, 2011; Schmitt, 2000). This can be seen in the researcher's observation of the RMCG and in the records of the vocabulary card movements made in every session. It is also proved by the results of the weekly immediate post-tests and the delayed post-test.

In conclusion, the RMCG had the opportunity to focus more on those words they wanted to learn, or on words that were more difficult for them to learn, as Nakata (2008) and Waring (2004) suggest. The RMCG classified difficult cards together and easy cards together to permit more revision of the difficult words and fewer revisions for easy items, as can be observed in

the word card movement analysis. The use of word cards in this study allowed for retrieval along with spaced learning, as suggested by Baddeley (1997).

2. *Performance in graded readers*

Incidental learning was evident from the GRG answers to the CQs on the graded readers. The GRG responses to the CQs showed that the participants paid attention to the target word and inferred its meaning from context. The GRG was learning target words without the intention of doing so, this corresponds with Richards and Schmidt's (2002) definition of incidental learning by focusing on the form and on the meaning. Guessing the meaning of a word from graded readers may also occur, according to Hulstijn (2003), Nation (2011) and Schmitt (2000), and this is part of the incidental vocabulary learning process. The GRG performance indicates that incidental vocabulary learning took place and the learners implicitly (i.e., automatically, naturally, unconsciously, and simply) learned by focusing on the meaning of language in the graded readers, and they were able to store forms unconsciously then use these forms automatically in the LDT and GFT later, as suggested by Hulstijn (2003) and Nation (2013).

Laufer (1996) claims that learners often come across words they do not know, or that they think they know but misinterpret, and this can result in these words hindering reading comprehension. Therefore, Laufer (1996) recommends the provision of clues in the graded readers, so the GRG may rely on guessing the meaning from context. This aims to avoid context redundancy and to make sure a target word appeared in different contexts because lacking a variety of context does not ensure the possibility of guessing new words. Also, it is advisable to limit the density of the unknown or the potentially misinterpreted words to help participants read smoothly. Efforts were made in the current research to achieve these goals.

The CQs were used in this study, although some researchers are against the explicit highlighting of target words. According to Waring and Takaki (2003) many researchers think that highlighting target words is a characteristic not found in natural texts and its use contradicts the purpose of graded readers. Some graded readers highlight target words just for the beginner levels (Waring & Takaki, 2003). However, the GRG benefited from the CQs as they could ignore the unfamiliar words to enjoy reading, or to avoid interrupting the flow of ideas while reading.

The GRG benefited from the number of repetitions of each target word (i.e., five times in two texts each week) in the graded readers. The higher the number of repetitions possible

increases the chance of learning unknown words, this confirms Pellicer-Sánchez and Schmitt (2010), Saragi et al. (1978), and Vidal's (2011) emphasis on the importance of repetition in graded readers.

There are many factors that affect word learning other than repetition, such as the ability to guess words from context and to make inferences (Laufer, 1996; Nation, 2014a; Nation & Wang Ming-tzu, 1999). These factors are addressed in Chapter 7 (see section 7.4.5 on limitations).

The GRG participants were not fluent readers of English, and therefore they might had struggled to remember ideas, as well as remembering each word involved, as suggested by Waring (1997). The Arab ESL learners in the GRG might had done what Waring (1997) suggests, which is to read each letter individually in order to decode the word and keep it in their WM. They might had experienced the problem to which Waring (1997) refers, i.e., forgetting the words kept in their WM when moving to the second line as they were at the intermediate level of language proficiency and they might had trouble connecting ideas in the text (Laufer, 1996). This can create frustration and may had slowed reading down for the GRG (Cobb (2007, 2008)).

3. *The efficiency of the LDT and GFT*

The LDT was designed to evaluate accuracy and the speed of word recognition by measuring the RT. Although the LDT instructions were clear, participants might press the response key to move on to the subsequent stimulus before they had fully evaluated their response to a target word. The participants were advised to be careful not to do this and were encouraged to give the correct response time as much as possible. The RMCG and the GRG seemed to take their time and to give responses to the best of their ability. However, the CG were less engaged in taking the test and this can be seen as they seemed occasionally to be pressing any key to move to the next stimulus without considering the accuracy of their responses.

The GFT was designed to measure productive vocabulary knowledge of target words. It was easy to explain the task to the participants, as completing the gap in a sentence with a specific word is a common drill used either for testing or for exercises on other courses in the English and Literature program at Albaha University. The sentences in the GFT provided both context and the first letters of the missing target word which helped the retrieval of the target word for the RMCG and the GRG. The design of the test was effective in terms of giving the

participants a clue of what exactly they should write in the gap in addition to the context provided in the sentence. The successful completion of most of the gaps by the RMCG and the GRG might be influenced by the Arab ESL learners' familiarity with this kind of test.

It is important to note that the repetition of the LDT and the GFT at the end of every week for three weeks and after two weeks from learning might had made the tests predictable. All three groups became familiar with some of the stimuli in the LDT and some of the sentences in the GFT. Target words from list A were repeated four times, target words from list B were repeated three times, and target words from list C were repeated two times. It was expected that this repetition would contribute to higher scores for the three groups for list A in the LDT, and for the RMCG and the GRG for list A in the GFT. Similarly, this repetition of test items caused the *testing effect* which contributed to the higher score recognition and production of list B than of list C. *Testing effect* might also explain why the CG, who did not receive any vocabulary guidance, showed an increase in recognition of target words from the first weekly immediate post-test to the delayed post-test. Finally, the LDT and GFT were effective for the purpose of the study. Moreover, testing of the same words many times might contribute to learning.

4. *Number of target words*

Fifteen target words seemed adequate for 20 minutes of learning a day for five days, based on the results of the Pilot study. This study was partly based on the Fitzpatrick, Al-Qarni, and Meara's (2008) study, in which one participant memorised 400 words for 20 days. However, such a small sample size is not representative of L2 learners. Furthermore, the subject was free to use the card file, to write notes and records, and had been encouraged to learn for 30 minutes, more or less, a day for the 20 days. Conversely, the study reported here controlled the time of learning for both the RMCG and the GRG at 20 minutes only, and the participants were not allowed to use the word cards and the *Leitner system,* or the graded readers, outside the classroom, or at a different time than during the learning sessions. Also, the participants were not allowed to take notes or study the target words in any other way than in the way they were guided.

Furthermore, the 45 target words are from the low frequency level (i.e., 5000 or more word-frequency level), unlike the words selected by Fitzpatrick et al., (2008) where the target words were from relatively high frequency Arabic words. Learning high frequency target words in the L2 is easier in comparison to learning low frequency target words. High frequency words can be encountered during the normal use of the language (Milton, 2009).

The number of target words was a manageable number for the RMCG to memorise in 20 minutes of learning over five days. As suggested by Mondria and Mondria-De Vries (1994), 15 target words is a manageable number of target words at the beginning of the learning experience as the number of cards first suggested for the first section is 3040 word cards. It is discouraging for learners to go through all 3040 cards in order to be able to move some of them to section two, therefore they suggest a fewer number of cards, namely 15 to 20, as a warm up.

Furthermore, it was important to have the same number of target words for both the RMCG and the GRG. The graded readers design, according to the suggestion made by Laufer (1989) and Nation (2001), ensured 95% coverage of familiar words in a text to develop an adequate level of comprehension. Also, the need to repeat target words ten times, as suggested by Webb (2007a; 2007b), makes any more than fifteen target words very difficult to implement in the graded readers taking into consideration the GRG limited time available for reading (i.e., 20 minutes), and the length of text they can read in 20 minutes according to their speed (i.e., 500- to 700-word text length).

6.3.1.3 Which of the two groups (the RMCG and the GRG) react fastest to the target-word as shown by the weekly immediate post-tests and the delayed post-tests?

The Hypothesis suggests that the RMCG would respond faster and would give a more predictable performance than the GRG. In fact, the RMCG responded faster to stimuli in the weekly immediate post-test LDT than the GRG. However, the GRG responded faster in the post-test LDT than the RMCG. Generally, the RTs decreased between LDT 1 and LDT 4 for the three groups, indicating a decrease in processing time across these conditions.

There was a significant difference between the RMCG and the GRG, and the RMCG and the CG in their first LDT RT but the effect size is medium for the difference between the RMCG and the GRG and small for the difference between the RMCG and the CG. There was also significant difference between the RMCG and the GRG and the GRG and the CG in their second LDT RT and the effect size is small in both cases. The three groups became marginally faster from LDT 1 to LDT 2, this could be for two reasons. The three groups were more cautious when they were taking LDT 1 because they needed to understand the task and, as they became more familiar with how to take the LDT, in the second week of learning they became faster. Furthermore, when the three groups took LDT 2, the stimulus set was larger and the participants found it more demanding to spend more time with each stimulus.

There was a significant difference between the RMCG and the CG and the GRG and the CG in their third LDT RT but the effect size is medium in both difference. The three groups became faster in LDT 3 for a number of reasons. First, the number of stimuli in LDT 3 was even larger than that for LDT 2 and was, therefore, more demanding for all three groups. Second, the three groups became familiar with the task between LDT 1 and LDT 3. Third, words in list A were tested twice in LDT 1 and LDT 2, and those in list B were tested once in LDT 2. The familiarity of the target words from lists A and B might have contributed to faster RTs.

After one week from learning, the RMCG and the GRG reacted to LDT 4 target words marginally faster than in the LDT 3. Although the RMCG and the GRG stopped learning for two weeks, they displayed high AS in LDT 4 and responded faster to target words than they did to target words in LDT 3. This shows a better level of recognition in which the RMCG and the GRG became more familiar with the target words.

The CG RT decreased from LDT 1 to LDT 4, but this did not necessarily indicate an increase in the processing time. Given that the CG had better AS and faster RTs the more they took the LDT, this may be a sign they were recognising some of the target words. In conclusion, the three groups became faster over time, indicating that the more the learners became familiar with the LDT the faster they performed. Furthermore, the longer the list of stimuli in the LDT, the faster the response to the target words.

6.3.1.4 How many forward and backward moves does the RMCG make in the Leitner system?

The results support the Hypothesis that the RMCG will use the *Leitner system* effectively and will succeed in moving word cards at least five times, so that they end up in the final section. Over the three weeks, the RMCG managed to space their learning and use the *Leitner system* to perform sufficient FWMs to help retain target words. The overall number of FWMs for the RMCG reached five, and the number of BWMs was very small. The number of FWMs for list A was marginally higher than for lists B and C, and the number of FWMs for list B was marginally higher than for list C.

According to Leitner (1972), when using the *Leitner system* the learner follows the process of moving flashcards forward until all the flashcards are moved to the fifth section, the end of this process means the learner has retained the information on the cards (as cited in Godwin-Jones, 2010). It was expected that FWMs between the lists would be significantly different due to the time of presentation and the length of time the lists remained in the Leitner

box. For example, it was expected that words from list A would be moved more than those in lists B and C. This is because list A was introduced in the first week of learning and its cards could remain in the Leitner box for two more weeks, but there was no significant difference between the FWMs of the three lists. However, the RMCG moved words from list A forward more than they moved others of the 45 target words, and these words were recognised and produced by the majority of the RMCG. For example, *proprietorship*, *elucidate*, *banish, pertinent,* and *vicissitude* were forwarded most frequently out of the 45 target word set, and the percentage of participants LDT AS and GFT scores for these words were relatively higher than for the rest of the 45 target words.

There is no indication that the fewer numbers of FWMs, or the greater number of FWMs, relate to higher recognition or production scores. For example, the words *shrewd* and *ingenious* from list A were the least forwarded by the RMCG and they have higher than average numbers of participant recognition and production. Yet, the words *ebullient* from list B and *astound* from list C had the highest FWMs out of the 45 target word set and show higher than average numbers of participant recognition in list B and in list C, respectively. In conclusion, the *Leitner system* is easy to apply and is an effective way to learn vocabulary intentionally. However, there is no clear indication that the more the learners move the word cards forward, the better the retention of the information.

6.3.1.5 How many comprehension questions does the GRG answer correctly by the end of each graded readers' texts?

The Hypothesis is that the GRG will be able to respond to most of the graded readers' CQs as the length of the texts and their difficulty were adjusted to participants' language proficiency. The GRG was successful in answering more than half the CQs, which is not very high, but it does indicate that the graded readers helped the GRG learn incidentally.

Pellicer-Sánchez and Schmitt (2010) argue that testing learners immediately after reading does not measure long-term learning as an immediate CQ test tends to draw attention to the target words, rather than enhancing long-term learning. Nation (2001) claims that learning from reading context might be low, except when the reader's attention was deliberately drawn to words in a specific context. The CQs were designed in a way that participants did not know they were being tested on their knowledge of the target words, instead they had to answer CQs that required their understanding of the meaning of the target words. Finally, although

reading graded readers with very similar characteristics to the natural texts, it is beneficial to drive the reader's attention to the target words indirectly by asking CQs.

6.3.1.6 Do Arab ESL learners' performance decrease when large quantities of vocabulary input are maintained over an extended and consistent period?

The Hypothesis is that Arab ESL learners' performance will decrease, to some extent, when large quantities of vocabulary input are maintained over an extended and consistent period. The results are compatible with this Hypothesis. Both the RMCG and the GRG performance decreased, to some extent, from one test to another. The RMCG was successful in recognising almost all the target words in the LDT by the end of the three weeks, the RMCG was also able to produce many target words out of the 45 target word set in the GFT. Although the RMCG was superior to the GRG in terms of word recognition and production, the GRG recognised and produced a considerable number of the target words after three weeks of learning. The significant difference between AS and GFT scores for all groups has a small size effect.

There are two main factors which contribute to the steady decline in performance of both groups. Target words fade from memory if re-learning does not occur to strengthen memorization. The forgetting of target words did not drop so sharply after re-learning sessions, as suggested by Baddeley (1997), Pimsleur (1967), Schacter (2002), and Waring (2004). For example, the RMCG recognition dropped from 100% to 87% in the first LDT AS, and from 87% to 85% in the second LDT AS; the drop did not exceed 13% for the RMCG over the four LDTs. In other words, recent exposure to a list played an important role in recognising and producing the words on it. The second factor is that the heavier the learning burden, the lower the recognition and production scores would be. Finally, learners might deal better with the learning burden in the early stages of learning. Also, the recent exposure to memory may contribute to better retention than the less recent exposure to target words. The following is a discussion of the results of the weekly immediate post-tests and the post-test for the RMCG and the GRG.

1. *The RMCG learning burden in the weekly immediate post-tests*

The RMCG had no difficulty in learning almost all the 15 target words in the first week

of learning. Rehearsal using the *Leitner system* and RM word cards was beneficial because the RMCG did not only recognize the form in the LDT, they were also able to produce the written form in the GFT. The learning burden was not yet heavy in the first week and this might have contributed to the high LDT AS and GFT scores in the first week.

The RMCG recognised a high number of target words in the second week. However, their learning burden increased in the second week. The percentage of participant recognition marginally declined from the first week to the second week of learning. The RMCG was also successful in producing target words in the second week of learning.

The successful recognition and production of target words in the second week of learning was due to the duration of exposure to the words on list A, which continued to be in the Leitner box for some of the RMCG participants in the second week of learning. Moreover, although the RMCG was introduced to list B, which has the same number of target words as list A (i.e., 15), and had the same duration of learning for the new list (i.e., one week) the learning burden became higher (i.e., from 15 to 30 words in total), and this contributed to the slightly lower LDT 2 AS for list B.

In the third week of learning, the RMCG recognised a high number of words out of the 45 target words. At this stage of learning, the RMCG were given the last addition of the target words and the learning burden became higher. RMCG recognition and production scores for lists A and B in the second week of learning declined in the third week. Although none of the list A words remained in the Leitner box in the third week for any of the participants, the RMCG recognised a high number of target words from list A. Furthermore, although most of target words from list B were not in the Leitner box in the third week for most of the participants, the RMCG retained most of the target words from list B. The RMCG also retained a high number of words from list C, so the recent exposure to list C might have contributed to their learning. Therefore, the learners perform better when the learning burden is lower, and recent exposure to the words contributes to better retention.

2. The RMCG learning burden in the delayed post-test

Although two weeks elapsed between the last immediate weekly post-test and the delayed post-test, the RMCG recognised and produced a high number of target words out of 45. The scores were higher for list A than for lists B and C, and the group recognised and produced higher numbers of words from list B than from list C. Although, for the RMCG, the elapsed time between the last exposure to list A and LDT 4 was four weeks, the highest recognition and production scores were for list A, indicating the lighter the learning burden the

higher the retention.

The RMCG recognised and produced more than half of list B, even with the long delay between the last exposure to list B and LDT 4 (i.e., three weeks). Moreover, despite the time gap between learning list C and taking LDT 4 and GFT 4 (i.e., two weeks), the RMCG successfully recognised more than half of list C and produced half of list C. The decline of the RMCG performance from list A to list B to list C in LDT 4 and GFT 4 can be attributed to the heavier learning burden created by the addition of lists B and C to list A. Finally, long delays between learning from the word cards and taking the post-test might not cause difficulty in retaining newly learned vocabulary when the learning burden is light.

3. *The GRG learning burden in the weekly immediate post-tests*

In a study conducted by Horst, Cobb, and Meara (1998), the 34 participants (from lower-intermediate language proficiency level) were introduced to longer graded readers' texts (109-pages) over 10 days. After comparing the results of a pre-test and a post-test, the results show the recognition of new vocabulary to be very low (i.e., one word in five was learned) (Horst et al., 1998). The GRG recognised and produced a high number of target words in the delayed post-test in comparison to previous studies. The GRG recognised 70% and produced 34% of the target words after two weeks from learning, while the learners in Waring and Takaki's (2003) study know the meaning of 6% of 25 new words after three months - no production tests were conducted.

However, the GRG had higher recognition and production rates for list A than for lists B and C in the weekly immediate post-test and the delayed post-test compared to the RMCG. The GRG retention of list A target words in the first week was due to the fact that the GRG received the LDT and GFT immediately after reading the graded readers with words from list A, so the target words from list A were fresh in the participants' memory. Moreover, the learning burden was relatively low in the first week.

In the second week of learning, the GRG recognised a high number of target words out of 30, but struggled with their production. Although the recognition in the second week was successful, the GRG performance for words in list A in the second week of learning declined. The decline in the GRG performance for list A words indicates that they struggled to produce them due to not reading graded readers containing list A items in the second week of learning. Furthermore, the GRG recognised a high number of list B items in the second week, but struggled to produce them in the GFT. The GRG read graded readers with list B words just before they took the second weekly immediate post-test, but recognition and production of list

B were not as high as for list A in the first week, indicating the impact of the higher learning burden on retention.

The GRG recognised a high number of target words out of 45 in the third week of learning. However, the GRG struggled to produce the 45 target words in the third week. The performance of the GRG for list A sharply declined in the third week of learning. This might be due to not being exposed to list A-related graded readers in the second and third week of learning, and also to the effect of the high learning burden by the third week.

There was a dramatic decline in the GRG performance for list B words in the third week of learning. This sharp decline was due to the gap in time between GRG exposure to list B and the test. Unlike list A, the sharp decline for list B was due to the higher learning burden the GRG experienced in the third week.

The GRG recognised a high number of target words from list C in third week of learning, but struggled to produce them. Although, the GRG was reading graded readers with list C target words just before they took the third LDT, the LDT AS and their GFT scores were lower than those for list A in the first week of learning. The factor of a lower learning burden in the first week than in the second and third weeks played an important role in the high LDT AS for list A words in comparison to the words in the other lists. The high burden of learning new target words might have contributed to this low production of list C words immediately after finishing reading the third graded reader. In conclusion, the lighter the learning burden from the graded readers the easier the learning will be. Also the more recent exposure to the new vocabulary contribute to better recognition and production.

4. ***The GRG learning burden in the delayed post-test***

The LDT AS in the post-test shows that the GRG recognition of lists A, B, and C were similar in LDT4. The similarity is despite the fact that the last exposure to list A was four weeks earlier, the last exposure to list B was three weeks earlier, and the final exposure to list C was two weeks earlier. This can be considered evidence for long term retention of target words, especially for items in lists A and B.

The time between GFT 3 and GFT 4 contributed to this decline in the GRG performance in the GFT. The production levels did not vary for lists A, B, and C in the post-test GFT. Despite the gap in time between the GRG exposure to lists A, B, and C, and the elapsed time before GFT 4, the GRG produced a very similar number of target words from each list. This also can be an evidence of long term retention of target words in the delayed post-test. Nevertheless, the GRG performed better with list A and this might be due to the fact that the

GRG did not have a high learning burden while learning list A words in the first week. The high burden of learning new target words might contribute to the low production of words from lists B and C after two weeks from learning. Finally, long delays between learning from the graded readers and taking the post-test might not cause difficulty in retaining newly learned vocabulary when the learning burden is light.

5. *The CG learning burden*

The CG performance fluctuated from the first weekly immediate tests to the post-test. The CG did not show high recognition rates for list A words in the first weekly immediate post-test. The performance of the CG improved a little more when list B was added to the second weekly immediate post-tests. The performance of the CG in LDT declined in the third week of learning, but it increased in the post-test.

As expected, most of the CG participants failed to produce target words in the GFT. The target words were from the low frequency level, according to the BNC, and the CG vocabulary size covered about 3000-word frequency level, according to the VST. Only one participant produced the word *proprietorship* and two others produced the word *shrewd* in the GFT 1. The CG production scores were not considered in the analysis because they were very low.

The CG performance per list shows that the CG was learning from tests. The CG recognition of list A words improved from LDT 1 to LDT 2. The recognition of list A words also increased in LDT 3. However, the CG performance in list B fluctuated. The CG recognition of list B words declined in LDT 3, then increased in the post-test. Furthermore, the CG recognition for list C words improved in the delayed post-test. The CG was not exposed to any of the target words in any way, yet their word recognition seemed to improve over time. This might be the effect of testing, or the repetition of taking the tests.

6. *Comparison of learning burden between groups*

As predicted, there was a significant difference between the RMCG and the GRG in word recognition and production in the weekly immediate post-tests. The RMCG outperformed the GRG in the post-test and the differences were significant. The RMCG shows the importance of repeating learning at different times so the newly learned information will be retained in the long-term, as suggested by the *Leitner system*.

The GRG performance declined, which confirms Pellicer-Sánchez and Schmitt (2010) claim that even with repetition in graded readers the amount of vocabulary retention from reading declines in the long term.

The CG had lower LDT AS and failed in the GFT. There was a significant difference between the CG performance and the RMCG and the GRG. Although the CG vocabulary size knowledge, language proficiency, WM capacity, and language contact were similar to the RMCG and the GRG (see the Methodology Chapter), there was a dramatic difference between the CG performance and the RMCG and GRG, and this demonstrates the efficiency of techniques used by the RMCG and the GRG. In conclusion, although vocabulary learning burden can hinder learning to some extent, intentional vocabulary learning using word cards is effective to ease the burden of learning an accumulative number of vocabulary over time. Incidental vocabulary learning using the graded readers might ease the difficulty of the vocabulary learning burden but it is not as effective as intentional learning from word cards. Also, not only learning techniques could help learners learn a heavier load of vocabulary items but also the frequency of exposure to target words via testing can play an important role in learning similarly to what happened with the CG.

6.3.1.7 How does the retention of target words in list A, list B and list C differ in the weekly immediate post-tests from the delayed post-test for the RMCG, the GRG, and the CG?

The results correspond with the Hypothesis for this sub-question, which is that learners in the RMCG and the GRG will perform better for list A than for lists B and C. In Appendix 6 (E) a summary of the results followed by discussion of the three groups' performance.

6.3.2 Do word characteristics affect word retention and learning for the RMCG and the GRG?

The results did not totally coincide with the Hypothesis that difficult word characteristics cause difficulty in target word retention for Arab ESL learners. The RMCG and the GRG obtained high scores for many of the target words with difficult characteristics. The response to this question consists of two parts. The first is about words characteristics that can

cause difficulty for ESL learners in general, specifically word length and derivational morpheme frequency. The second part is about special word characteristics that may cause difficulty for Arab ESL learners specifically, such as difficult phonological sounds, vowel blindness, and the translation of target words into Arabic.

6.3.2.1 How do word length and derivational morphemes frequency affect word retention for Arab ESL learners?

The Hypothesis suggests that longer words and words with low frequency derivational morphemes will be harder to learn and to retain by Arab ESL learners in the RMCG and the GRG. As seen in the Methodology Chapter, the target words were selected to have different lengths and different derivational endings with three levels of frequency.

The analysis of the recognition and production of target words with these characteristics shows they affected Arab ESL learners to some extent. The following is a discussion of the qualitative analysis classified into three sections: word length effect on word retention; frequency of derivational morpheme effect on word retention; and, list difficulty.

1. ***Word length: Long words***

It was expected that long words would be harder to retain than short words; however, the RMCG and the GRC obtained high scores for many long target words. For example, the word *new-fangled* had the highest percentage of participant recognition in LDT AS for both the RMCG and the GRG. Unexpectedly, the highest number of participant recognition in the CG was also for the word *new-fangled*. Not only was the word *new-fangled* recognised reliably by the three groups, it was also produced frequently by the RMCG and the GRG. The observation was made that some participants from the RMCG and the GRG reported that the word *new-fangled* was the easiest to learn because it started with the word *new.*

Many other long words were recognised reliably by the three groups and were also produced frequently by the RMCG and the GRG. This provides evidence that long target words can be retained in long term-memory when using word cards in conjunction with the *Leitner system* or the graded readers. Moreover, the RMCG recognised more long words than did the

GRG indicating the efficiency of intentional learning from word cards in conjunction with the *Leitner system* over incidental learning from graded readers.

It seems that some long words were more difficult than others. The RMCG and the GRG obtained lower AS and GFT scores than average for a number of long target words. It was not clear if the length of the words caused the difficulty, but this factor might have contributed to the difficulty.

2. *Word length: Short words*

As expected, a large proportion of the RMCG and the GRG recognised and produced most of the short words from the 45 target words list. For example, the word *shrewd* was recognised reliably by the RMCG and the GRG and it was also produced frequently. Some of the short words were reported as easy to learn and this coincides with the higher than average recognition and production rates. For example, *trespass* was reported by several participants in the RMCG as easy to remember because of the "unique sound" of it, i.e., it has the sound /s/ in two of its syllables. Several participants in the GRG noticed this word too, and asked about its meaning while they were reading.

Some short words, such as *toil,* were difficult for both the RMCG and the GRG. Although *toil* is a one syllable word, it had very low numbers of participant recognition in the LDT and production in the GFT for both the RMCG and the GRG. It is not clear if this word was difficult because of its meaning, or because it appeared in the last list.

3. *Morphological Aspects of the Target Word*

It is not clear if the frequency of the derivational morphemes really affected the RMCG and the GRG recognition and production. Some target words with no derivational morpheme were highly recognised and produced by the RMCG and the GRG (e.g., *trespass* and *shrewd*). Yet, some target words with no derivational morpheme were not highly recognised or produced (e.g., *toil*).

Moreover, it was not clear if high frequency derivational morphemes in a target word contributed to learning. For example, *new-fangled* has the derivational morpheme *-ed* and was recognised reliably by the two groups, it was also produced frequently. However, although

afflicted has the same derivational morpheme -*ed* (when used adjectivally), it had lower than average numbers with respect to the participant recognition and production for the three groups.

It was also not clear if target words with a low frequency derivational morpheme hindered learning. For example, *proprietorship* had the least frequent derivational morpheme (-*ship*) of those in the target set, yet it was recognised reliably by both groups and it was also produced frequently. It seems that learning via word cards in conjunction with the *Leitner system* was effective to learn what was considered to be difficult. *Vicissitude* was reported as 'hard to learn' by several RMCG participants and it contains the low frequency derivational morpheme -*itude.* The results for this target word were striking as the RMCG and the GRG were able to highly recognise and produce it despite the fact that the RMCG commented on its difficulty, and despite the presence of a low frequency derivational morpheme.

4. *List difficulty*

According to Dosher (2005), words with phonologically different features are easier to remember than words with similar phonological features. The very low recognition and production of *cogent* may be an effect of its similarity to *concur, contrive, contentious,* and *colossal.* However, *concur, contrive,* and *contentious* were highly recognised and produced by the RMCG and this might be because of the efficiency of repeating them in the *Leitner system*. The RMCG participants reported their struggle with *contentious* while they were learning, yet the RMCG highly recognised and produced it.

Unlike learning from a list written in one place, learning from word cards allows learners to keep changing the sequence of words and classify them according to their needs (Nakata, 2008). Learning from word cards overcome the drawbacks of learning from word lists such as not being able to pay enough attention to each target word even with similar phonological sounds. The length of a list written in one place has an influence on the learner's ability to memorise words; long lists are more difficult to recall than shorter ones (Lewandowsky, 2005). Yet, the RMCG retained high numbers of the 45 target words, both in the short and the long term. The RMCG avoided what Nakata (2008) calls the *list effect*. Learners might remember some words while memorizing a list, but they cannot remember them when using these words in real life.

For some of the target words, such as *perceptive, perseverance, apprehensive,* and *pertinent,* the RMCG and the GRG had difficulty in recognising and producing them. Therefore, the RMCG might be affected to some extent by the similarity of phonological

features of some of the target words, even when able to classify words and move them according to their own needs.

The GRG had more difficulty with words with similar sounds than did the RMCG. For example, *concur, cogent,* and *contentious* were more difficult for the GRG than for the RMCG, along with other target words with similar phonological features. It was expected that the GRG would have more difficulty in recognising and producing target words with very similar phonological features as they were learning target words incidentally. The better recognition and production by the RMCG of the target words with similar phonological features highlights the advantage of intentional vocabulary learning over incidental vocabulary learning.

6.3.2.1.1 How do phonological characteristics affect word retention for Arab ESL learners?

The results partly coincide with the Hypothesis that target words with difficult phonetic features for Arab ESL learners will be hard to learn and retain. As mentioned in the Literature Chapter, Arab ESL learners have special difficulties with word recognition. Fender (2008) states that Arab ESL learners have more difficulties in reading and literacy. Also, he states that Arab ESL learners struggle with English spelling in particular, and this affects their reading skills. According to Saigh and Schmitt (2012), struggling with English spelling affects their writing skills. Fender (2008) states that there is a relationship between phonological and orthographic decoding of words and word recognition. The Arab ESL learners ability to decode phonological and orthographic aspects of English words affects their ESL ability to read, to acquire vocabulary, and to pronounce words proficiently, therefore these skills affect word recognition (Fender, 2008).

The qualitative analysis reported that sometimes the special phonological difficulty and translation to the first language contributed to low retention rates of some target words. The comments from the participants about some target words emphasized the results of the analysis. The following is a discussion of the qualitative analysis and it is classified into two sections: difficult phonological aspects of the target words; and, the effect of translating into Arabic.

1. *Difficult phonological aspects of the target word*

Arab ESL learners transfer their phonological knowledge of the L1 to the L2 and they use the way they write the sounds in spelling in their L1 in the L2 (Saigh & Schmitt, 2012). As a result, this transfer from Arabic can improve learning the L2 if the transfer is from similar phonological/orthographic systems, or can cause problems if the transfer is from different

phonological/orthographic systems (Saigh & Schmitt, 2012). Arab ESL learners transfer L1 phonological systems to a greater extent in comparison to other ESL learners (Fender, 2003). The following is a discussion of the Arab ESL learners' performances of target words with consonants difficult for Arab ESL learners, a discussion of Arab ESL learners' performance of words with consonant clusters, and a discussion of the vowel blindness phenomenon that might affect Arab ESL learners in the current study.

a. Consonants Difficult for Arab ESL learners

There was no clear indication if Arab ESL learners in this study transferred their phonological knowledge of their L1 to the L2, or if they used the way they write the sounds in spelling in their L1 in the L2, as suggested by Saigh and Schmitt (2012). Arab ESL learners find difficulty with sounds in English that do not exist in Arabic. The comparison made is for the words with the sounds /p/, /v/, and /g/. It seems that both groups had high numbers of recognition and production of the words with these difficult consonants, yet they also had difficulty in recognising and producing other words with the same difficult consonants. For example, target words with the sound /p/, such as *trespass, proprietorship,* and *perceptive,* were easily recognised and produced by the RMCG and the GRG. However, there were other target words containing the sound /p/, such as *perseverance* and *petrify,* which were not easily recognised or produced by either group.

b. Consonant Clusters

Another difficulty that Arab ESL learners struggle with is initial consonant clusters in English. There were no target words with cccv, or cccvcc format in the target word list. Only one target word appeared which consists of two initial consonant clusters and it was not difficult for the RMCG or the GRG. This word was *proprietorship* and it had high numbers of participant recognition and production by both the RMCG and the GRG.

c. Vowel Blindness

Saigh and Schmitt (2012) claim that Arab ESL learners show a certain difficulty in recognizing short English vowels in a phenomenon called *vowel blindness*. In *vowel blindness* Arab ESL learners do not know the correct vowel to use and this causes difficulties with word recognition (Saigh & Schmitt, 2012). It seems that the RMCG performed better than the GRG for target words with short vowels. The difference was striking between the RMCG and the GRG in recognising target words with short vowels, such as *mystification* /mɪstɪfɪˈkeɪʃən/ with

the short vowel /ɪ/ occurring three times, and the target word *idiosyncratic* /ˌɪdɪəʊsɪŋˈkratɪk/ also with the short vowel /ɪ/ occurring three times.

It seems that target words with long vowels were easier to retain for both the RMCG and the GRG, indicating that Arab ESL learners provide evidence of *vowel blindness.* The RMCG and the GRG showed better recognition and production of the target words with the long vowel /uː/, such as *stupendous* /stjuːˈpɛndəs/ and *vicissitude* /vɪˈsɪsɪtjuːd/. However, they showed low recognition and production of most target words with the long vowel /iː/, such as *feeble* /ˈfiːbl/, *tedious /ˈtiːdɪəs/,* and *grievous* /ˈgriːvəs/.

2. The effect of translating into Arabic

On the one hand, target words with familiar meaning and frequent usage in Arabic, such as *shrewd,* were much easier to retain by the RMCG and the GRG. On the other hand, the more the translation of a meaning was abstract and unfamiliar in Arabic, such as *pertinent,* the more difficult the word was to recognise or produce. Although it was difficult for Arab ESL learners to learn target words with abstract meanings, it seems that the RMCG overcame this difficulty for some of the target words, e.g., *perceptive.* Based on the comments made by the RMCG, they reported words with abstract meaning as difficult, but they successfully retained these words indicating the efficiency of intentional learning using word cards.

6.3.3 To what extent does repetition and spacing affect learning using word cards and graded readers?

The alternative Hypothesis for this research question is that repetition, spacing, and time of learning will affect learning in both groups**.** The results partly coincided with this hypothesis by showing how important it is to employ spaced repetition using word cards to achieve the greatest retention of information, as suggested by Godwin-Jones (2010). As illustrated by Mondria and Mondria-De Vries (1994), moving word cards forward indicates that the learner is deciding that this card is learned, and moving word cards backwards indicates that the learner is deciding that this card is not learned and needs further revision. The partial correlation of FWMs with LDT AS and GFT scores, and the partial correlation of BWMs with LDT AS and GFT scores, showed how repetition and spacing using word cards affected learning of the target words.

For every target word in the graded readers there was a CQ. Therefore, the CQ scores could be an indicator of the effects of repetition and spacing. The CQs correct responses were indicators of the retention of target words from context. In the Results Chapter, partial correlation of CQ scores for the GRG with LDT AS and GFT scores were also analysed to show to what extent repetition and spacing affected learning using graded readers. The following is a discussion of the impact of time and number of repetitions on word learning and retention, followed by discussion of the impact of spacing and time of learning on the learning and retention of target words.

6.3.3.1 What impact does the time and number of repetitions of the 45 target words, 1 to 8 times for the RMCG and 10 times for the GRG, have on word learning and retention?

The results partially coincided with the Hypothesis that target words need to be repeated at least five times for the RMCG to learn and retain them. Words need to occur five times in each of the graded readers and twice in five learning sessions before they are learned and retained. Vidal (2011) claims that there is a correlation between repetition and vocabulary learning from graded readers, most of the time. However, in the current study the number of repetitions did not correlate significantly with the GRG scores in the LDT and in the GFT. It is true, however, that repetition affected learning for the GRG as the most frequently repeated list (i.e., list A) was learned better than the less frequently repeated lists (i.e., list B and list C). Lists A and B did not only appear in the graded readers but also in the tests, and this may have positively affected the learning and retention of these words. There was only one significant partial correlation of FWMs with the GFT scores for list A in the first week indicating that repetition of the target words using word cards affected target word production.

The impact of spaced repetition using word cards and the Leitner system

As expected, there was a significant positive correlation between the percentage of participants who managed to learn target words from list A in the first week and the percentage of participant recognition of the target words in the first LDT, and the numbers of participant production of the target words in the first GFT also. Moreover, there was expected to be a significant positive correlation between the percentage of participants who managed to move

target words from list C to the fifth section in the third week and the numbers of participant production of the target words from list C in the third GFT. These positive correlations indicate the efficiency of spacing using word cards and the *Leitner system.*

Furthermore, the positive and significant correlation between the percentage of participants who managed to learn target words from lists A, B, and C in the three weeks and the numbers of participant recognition of the target words in LDT 4 emphasises the efficiency of spacing using word cards and The *Leitner system*. There was a positive and significant correlation between the percentage of participants who managed to learn target words from lists A, B, and C in the three weeks and the numbers of participant production of the target words from lists A, B, and C in GFT 4.

However, unexpectedly, there was no relationship between the percentage of participants who managed to learn target words from list B in the second week and the numbers of participant recognition of the target words in the second LDT. Moreover, there was no significant relationship between the percentage of participants who managed to learn target words from list B in the second week and the numbers of participant production of the target words in the second GFT. Also, unexpectedly, there was no significant relationship between the percentage of participants who managed to learn target words from list C in the third week and the numbers of participant recognition of the target words in the third LDT.

a. Easy words

As predicted, the easiest target words to learn were the words that were moved to the fifth section by most of the participants as they were the most recognised and produced words by the end of the first week. Words that were repeated five times or more were such as the words *proprietorship*, *shrewd,* and *ingenious*, most of the time, were easier than target words with fewer repetitions.

b. Difficult words

According to expectations, words that were not moved to the fifth section by most of the participants by the end of the end of the week in which they were introduced had lower numbers of participant recognition in the LDT and production the GFT by the end of the first week, these words include *rambunctious* and *expeditious.*

6.3.3.2 Do time of learning and spacing affect learning from word cards and graded readers?

The impact of space between the occurrence of the target word in the graded readers on the GRG performance

The results show that target words with all types of spacing had very close numbers of participant recognition in the immediate LDT and post-test LDT, and participant production in the immediate GFT and the post-test GFT. Therefore, spacing between the occurrence of target words did not affect their retention.

The results show that the time of target word occurrence in the graded readers had an impact on learning. This result shows that the fresher the target words were in the participant's memory, the more chance there was of their retention. This confirms the forgetting curve theory which suggests that with no re-learning a sharp drop at the outset is followed by a gradual decline if no re-learning occurs (Baddeley, 1997; Schacter, 2002).

The multiple regression analysis shows that the frequency of exposure to target words predicted LDT AS and GFT scores for the RMCG. Moreover, the multiple regression analysis showed that the number of syllables in a target word predicted the GFT scores of the GRG and the GFT scores for all groups combined, i.e., longer words were more difficult for learners than shorter words.

6.4 Conclusion

The results coincide with what was illustrated in the Literature Review in Chapter 2, that intentional vocabulary learning using the word cards in conjunction with the *Leitner system* resulted in better vocabulary retention than incidental vocabulary learning using the graded readers. The results also coincide with what was illustrated in the Literature Review in Chapter 2 about repetition and re-learning as they are important and are more effective when they occur intentionally from the word cards in conjunction with the *Leitner system*. The RMCG production scores were related to the frequency of exposure to the target word. This indicates that repetition and re-learning were more effective for the RMCG than for the GRG.

Furthermore, the RMCG were able deal better with the learning burden better than GRG and the CG and they were better in retaining more target words with difficult characteristic (i.e., long words, words with low frequency derivational morpheme, word with difficult consonants and words with short vowels).

7 Conclusion

7.1 Introduction

According to Lewis (1993), “lexis is the core or heart of language” (p. 89). This study adds to vocabulary research, vocabulary teaching, and vocabulary learning in general, and adds to the research on intentional and incidental vocabulary learning using learning techniques specifically. Consideration of the pedagogic and personal factors, and the context of the study discussed in the Introduction Chapter has contributed to designing and choosing the aims of this study.

The current research revealed that intentional vocabulary learning using word cards in conjunction with the *Leitner system* resulted in better and faster word retention for Arab ESL learners at the intermediate language proficiency level compared with incidental vocabulary learning using graded readers. Arab ESL learners who used the intentional learning technique dealt better with the learning burden than did Arab ESL learners who used the incidental learning technique. In addition, the qualitative analysis revealed that Arab ESL learners who used intentional learning techniques dealt better with common difficult characteristics and target words, and with special difficult characteristics than Arab ESL learners who used incidental learning technique.

Timing and spacing of learning the target words contributed to the ease or difficulty of learning for the RMCG. However, there was no clear indication that spacing in the GRG played an important role in the recognition of target words in the form recognition and usage production tasks.

Finally, the Discussion Chapter provided supporting evidence and a description of the results and showed the degree to which this study has achieved its aims. The Conclusion Chapter addresses the limitations and provides an evaluation of the current study, its pedagogical implications, and recommendations for future research.

7.2 Implications for Vocabulary Learning

Although this study had been conducted with Arab ESL learners at Albaha University in Saudi Arabia, there are a number of generalisations that can be made. Firstly, learning vocabulary intentionally using word cards in conjunction with the *Leitner system* is more effective than learning target words from incidental learning using graded readers.

For immediate and long term learning the word cards in conjunction with the *Leitner system* are more effective for the retention of a considerable number of vocabulary items. Learners with low to intermediate language proficiency, and with low vocabulary size, can benefit more from intentional learning using word cards and the *Leitner system* than from incidental learning using graded readers because they can learn the form and meaning of a large number of target words in a relatively short time. Furthermore, in conditions similar to the current study, where learners have little exposure to the L2 outside the classroom, learners need to focus on learning vocabulary intentionally because they have little real contact with the L2

Intentional vocabulary learning using word cards and the *Leitner system* can lead to high numbers of target word recognition and production. Intentional learning using word cards can be successful when paying attention to the design of the word cards (i.e., free design vs ready-made), and when word cards are utilized in the *Leitner system*.

For successful intentional vocabulary learning using word cards there should be flexibility in the memorizing of new vocabulary using word cards. Making decisions about the need to revise the target words, or not, by the learner is effective in retaining vocabulary. Knowing how and when to move cards in the *Leitner system* is important as there was a significant correlation between FWMs of cards to fifth section in the *Leitner system* and recognition and production of the target words.

Graded readers helped the Arab ESL learners learn and recognise more than half the list and produce one third of the target words incidentally. Learning vocabulary incidentally can lead to recognition and production of more than half of 45 target words. Drawing the beginner learners' attention to target words while reading graded readers was effective. The CQs played an important role in helping the learners to retain more than half of the 45 target words.

However, as claimed by Nation and Wang Ming-tzu, (1999) graded readers are the only way that beginner second language learners can enjoy reading a L2 with ease as real or natural reading texts are very difficult for L2 learners who are at a low proficiency level (Nation & Wang Ming-tzu, 1999). Observing the GRG while reading the graded readers confirms that

graded readers should be used in the classroom to ensure teachers' individual observation, as suggested by Waring (1997).

Fifteen target words each week was suitable for intentional and incidental vocabulary learning. Naturally, when learners were dealing with large quantities of vocabulary input over an extended and consistent period, their performance declined gradually. However, retaining more than half the list in the long term was staggering. Therefore, it is important to think about not only the number of target words the learners can retain but also for how long they can retain them.

The current study shows how the *testing effect* can result in learning. The CG did not receive any vocabulary guidance of any kind. Yet, the CG word recognition improved from LDT 1 to LDT 4, indicating that taking the test several times might contributed to the improvement of performance. The CG might have experienced the *retrieval practice,* or what is commonly known as the *testing effect* (Goossens et al., 2016; Roediger & Butler, 2011; Rowland, 2014) in which they retained some of the target words just from taking the test a number of times.

Finally, the time of learning of target words might have contributed to the ease or difficulty of learning target words intentionally and incidentally. Spacing of target words contributed to the ease or difficulty of learning target words intentionally and incidentally. Repetition of target words contributed to the ease or difficulty of learning target words intentionally and incidentally. Spaced learning via intentional learning using word cards and the *Leitner system*, or incidental learning using graded readers, helped the Arab ESL learners learn and retain most of the 45 target words.

7.3 Pedagogical Implications

The first pedagogic implication from the results of this study is the importance of explicitly drawing attention to certain features of vocabulary by using word cards and the *Leitner system,* and the careful design of incidental vocabulary learning via graded readers. Furthermore, based on Nation's (2013) suggestion, and from the results of this study, it is effective to meet vocabulary items in different ways to gain a range of aspects of knowledge of a vocabulary item. Specifically, it is important to help learners know ways of learning vocabulary intentionally, and to provide ways to help learners meet with the same vocabulary incidentally.

This study challenged the assumption some teachers hold about explicit vocabulary teaching and intentional vocabulary learning by over-evaluating implicit vocabulary learning, as claimed by Nation (2013). The RMCG outperformed the GRG in the weekly immediate post-tests and the delayed post-test and was able to deal with the learning better than the GRG. Moreover, the RMCG long term retention of target words was better than the GRG.

This study provides practical research evidence that can simplify the job of teaching and learning vocabulary. Teachers and learners can benefit from the economical and practical use of the *Leitner system*. This study also provides evidence that low frequency words can be learned intentionally by using word cards and the *Leitner system*. Moreover, teachers might use the idea of designing the graded readers and integrating them into the classroom with intentional learning from word cards and the *Leitner system*.

Teachers and learners can benefit from paced repetition using word cards. As mentioned in the Introduction Chapter, word card techniques ensure student-centred principles, not only because they can be used independently but also because the learners can personalize their cards according to their needs and desires (Elgort, 2011). Although this study showed the downside of the free-design word cards, this was mainly because the learners have to have similar experience of the target words (i.e., participants had to write down the same information), and because time did not allow for more experimenting with the free-design word cards.

Teachers can train learners to use the *Leitner system* to maximize the benefits of word cards, as suggested by Mondria and Mondria-De Vries (1994). The *Leitner system* applies the spaced repetition technique and this can be manually applied by using word cards and a shoe box, as suggested by Mondria and Mondria-De Vries (1994). The teachers can keep track of the learners' performances or can provide guidance, such as when to add more word cards, and what target words the learners need to learn according to their vocabulary knowledge, and they can test their knowledge of these target words when they reach the fifth section.

This study also showed the Arab ESL learners vocabulary growth rate from week 1 to week 2 to week 3 and to the delayed post-test after two weeks from learning intentionally and from learning incidentally. As mentioned in the Introduction Chapter, according to Nation (2013), vocabulary research is important not only to provide information of the suitable techniques to teach and learn vocabulary but also to focus more on how vocabulary knowledge grows over time.

7.4 Limitations and Possible Criticism of the Study

While a number of limitations of the study have been acknowledged elsewhere in the thesis, this section aims to provide a summary of those and to rationalise their effect on the study as a whole. First, general limitations are addressed, followed by limitations of the pre-tests. Then, the limitation of using word cards and *The Leitner system,* and the limitations of the graded readers design are addressed. Following this is discussion of the limitations of the LDT and the GFT.

7.4.1 General limitations

Limitations of the current study involve the number of participants and the shortage of time and resources. There are relatively few participants and a much larger number is more desirable so the effect of intentional learning from word cards and incidental learning from graded readers could be seen more clearly. The number of target words (i.e., 45) might be too limited. Increasing the number of target words would test the limit to which learning the form and meaning of target words in three weeks can be achieved. However, time did not permit such an investigation. Furthermore, the effect of intentional learning from word cards and the *Leitner system* and incidental learning from graded readers would be more informative if the delayed post-test interval was extended. For example, testing the vocabulary retention after one or two months would show the extent to which learning intentionally from word cards and incidentally from graded readers is effective in the long term. Moreover, the limited time available for this study affected the design of the weekly immediate post-tests and the delayed post-test.

There is a slight imbalance in comparability of time on task across CG group and the RMCG and the GRG. The CG received no training to help promote their vocabulary learning while the RMCG and the GRG were both learning vocabulary in a 20-minute per session. It would be beneficial to consider exposing the CG to a vocabulary learning technique but it was not possible as the purpose of having the CG is to compare to what extend learning using the word card and *Leitner system* and learning using the graded readers are different from not receiving any vocabulary guidance. Nevertheless, the CG showed a slight progress in learning vocabulary over time and this is like to be due to the multiple testing they received.

Furthermore, limited time and resources did not allow the investigation of the advantages of computer-based word cards over the paper-based word cards. Nakata (2008) compared computer-based word cards with paper-based word cards and word lists. The

computer-based word card group significantly out-performed the paper-based word card group and the word list group, and learners showed their preference for PC techniques over others. Due to the limited time and resources the investigation of computer-based word cards was not feasible.

7.4.2 The concept of intentional and incidental learning

Although it appears that incidental and intentional vocabulary learning are two different concepts, some researchers argue that they are two poles on the same continuum (Hulstijn, 2003; Nation, 2013). Learners can move between pure incidental and pure intentional learning, and vice versa, in the process of vocabulary learning (Hulstijn, 2003). This incidental-intentional continuum shows that integrating both types of vocabulary learning should occur at different times of learning with the focus on intentional learning or incidental learning according to language proficiency levels (Nation, 2013). In this study, intentional vocabulary learning was not integrated with incidental vocabulary learning and this integration might show the ultimate benefit of using two powerful ways of learning vocabulary.

7.4.3 Limitations of the pre-tests

The VST is a test created by Nation and Beglar (2007) and is one of the vocabulary profiles tests that measure learners' knowledge of words from high to low frequency profiles according to the BNC. According to Morris and Cobb (2004), vocabulary profiles tests are economical, easy to run, and can be done with limited time and resources. Schmitt (2014) agrees with these advantages, but he urges researchers not to misuse the VST. Schmitt (2014) argues that the VST does not measure the learners' ability to use the target words in reading, writing, listening, and speaking. In other words, answering an item correctly does not equate to the ability to use the target word. Schmitt (2014) also argues that the VST does not measure the depth of target word knowledge. Moreover, Kremmel and Schmitt (2016) claim that no matter the number of target vocabulary the learner knows in the VST, it is meaningless if they cannot use it in their language. These criticisms were considered and the VST was used in this study for two reasons: firstly, to make sure the participants share similar or very similar vocabulary knowledge; and secondly, to have an indication of the vocabulary frequency band with which the participants are more familiar. Mostly, the participants were more familiar with the 3000-word frequency band and this helped in the graded readers design and the selection of the target

words. The VST in this study was not meant to measure the participants' ability to use vocabulary in the four skills or to measure the participants' depth of vocabulary knowledge.

7.4.4 Limitations of word cards and the *Leitner system*

One limitation of the previous research may be that no study has empirically investigated the relative effectiveness of word cards and the *Leitner system* and the graded readers. Although researchers have frequently claimed the superiority of cards over other vocabulary learning techniques, or the advantage of spaced repetition using word cards and the *Leitner system* over learning word cards only with no specific plan, such arguments lack empirical support and remain speculative.

In addition to the low scores of the free design word card group in the pilot study, it was difficult to ask the participants to design their own cards. There was limited time for the participants to design cards for 45 target words, at the same time having to revise the accuracy and appropriateness of the cards for 25 participants. The cards have to be similar for all participants and the information written on the cards has to be correct. For example, it was intended that participants would learn the verb part of speech for the word *trespass* with the synonym *infringe.* The participants might choose to write the noun part of speech of the word *trespass* with the synonym *offence.* This problem was observed while conducting the pilot study and the participants were asked to write down the same definition and part of speech to avoid differences between participants in the word card design. Being guided in what exactly to write is not what is originally intended for free-design cards.

In addition to limitations of the free word card design is the limitation of the number of word cards used in *Leitner system*. Maintaining 15 word cards in the Leitner box for most of the week of learning was not possible without the close observation of the researcher, nor without continuous encouragement to take time while studying the cards. During the rehearsal it was important to keep reminding the participants that they should be very careful about their decisions to move the cards, and they should only move cards when they remember all the information written on the reverse side. Furthermore, they were continuously reminded to know the sound and the spelling of the target words. Without this monitoring and the continuous reminding of the how the task should be completed, the participants would move the cards in a speedy way and they would finish studying the fifteen cards within one or two learning sessions. The limited time allowed for the session (i.e., 20 minutes) was suitable only with the participants' careful movements of cards.

Therefore, it is recommended that the card number be higher if the learner does not want to spend only 20 minutes learning, or the amount of information on the reverse side of the card

be reduced. When the number is higher, the interval between the first meeting with a word card and the second meeting with the word card will be longer. For example, if the number of word cards is 100, the learners have to shuffle through 99 word cards before meeting the word card again, and this will allow for a longer interval between the first and second exposure. Longer intervals are more beneficial to vocabulary retention than shorter intervals.

Although paper-based words and the manual movement of them are advantageous, there are a number of limitations that should be considered. With the paper-based word cards used in the current study, there is no fast and easy access to resources or to other learners. Additionally, there is not the advantage of immediate feedback on success and progress, or activities that require participants to listen to the pronunciation of vocabulary items, or that require learners to record their voices as there is when using computer-based word cards. Paper-based word cards also lack the advantage of being able to save a record of a learner's progress in learning every individual word acquired, or to plan the sequence of learning, which is very useful even for learners with limited skills, as suggested by N.C. Ellis (1995), Nakata (2008);(2011) and Nation (2013). Using paper-based word cards in the *Leitner system* is somehow more demanding because learners have to make a number of decisions in order to classify their cards, as suggested by Nakata (2008). Paying attention to learning would interfere with planning a strategy for distributing word cards, and time spent in reviewing can result in an emphasis on learning already known words and a tendency to avoid learning difficult words (Nakata, 2008).

Finally, due to the need for smooth and uninterrupted learning sessions and to the difficulty of reporting all word movements for 45-word cards for fifteen learners during a 20 minute learning session, the word card movements during the sessions are not recorded. It was more practical to record the number of the last box in which each word card is found at the end of each learning session. Unfortunately, not being able to see the exact movement of word cards within the session does not provide the exact FWMs and BWMs. It would be beneficial if there is another way of recording the word card movement while learning without interrupting the learning process. This can be accomplished if the researcher is able to observe every learner individually and video record their card movements.

7.4.5 Limitations of the graded readers design

It would be more effective to test on much longer texts. The graded readers were modified to match the lowest threshold possible, which is 5000, as suggested by Laufer (1996). However, the learners knew less than 5000. The participant vocabulary size in the VST ranged from the 3000- to the 4000-word frequency. The words in the graded readers from the 5000

band could have been changed to words from higher frequency levels, but this was difficult due to time constraints.

Furthermore, native speakers were asked to proofread the texts, but it was difficult to make these low frequency words meaningful in the context. Low frequency words are rarely used because they are more specific to certain topics and to use them appropriately in the graded readers was challenging. All in all, the majority of the GRG performed well in the CQs, the weekly tests, and the post-test, thus the conditions for successful learning from context seem to have been met and comprehension of the text was achievable.

It is important to note that the graded readers design was not easy and was not fully successful. The graded readers had to be modified in a very limited time. The VST results were analysed a week before the experiment and, as the results show, due to the vocabulary size of the learners it was challenging to simplify long texts in a very short time. Native speakers were asked to help edit and proofread the texts, but more time and work was needed. Moreover, the modified graded readers used for the current study were perhaps not entirely successful because the topics were not sufficiently engaging for the readers.

In addition to the limitations in the graded readers design mentioned previously, it would be more efficient to investigate factors such as: the learners' mastery of reading skills in the L1 and the L2 e.g., guessing words from context, making inferences, predicting and comprehending the meaning of the paragraph; and the learners' ability to use their experiences and background knowledge to understand the text.

Due to the limited time available for this study, and the focus on incidental vocabulary learning, it was not the aim to investigate the GRGs' ability to guess words from context, to make inferences, or to predict and comprehend the meaning of the paragraph. Moreover, even if there are factors other than the vocabulary that are important for understanding a text, the threshold of reading comprehension is mainly lexical (Laufer 1996). However, while choosing the graded readers to be modified it was taken into account the learners' ability to use their experiences and background knowledge to understand the text, and the learners' tendency to be engaged with the content of the graded readers (Laufer, 1996; Nation & Wang Ming-tzu, 1999).

Although the GRG were able to recognise and produce a considerable number of target words, the reason might not necessarily be related to their reading skills. There was a sense that the GRG knew that they would be tested on information in the CQs. They were not told this, but from knowing the educational culture in this context it was predictable that they would guess so.

It is not possible to avoid the CQs in the graded readers design. Focusing on the CQs prevented the GRG from reading the text without attention to the meaning and just skimming or scanning the text to find the answers. The CQs helped to ensure their engagement with the

text and the high scores in the CQs might be the reason for their reasonably good performance on the LDT and the GFT.

Another limitation of the graded readers design is the difficulty to get the GRG involved in silent reading as the graded readers are time-consuming, as suggested by Waring (1997), and this results in less enthusiasm among the GRG than the RMCG to do the task.

7.4.6 Limitations of the LDT and the GFT

It was challenging to design four LDTs and four GFTs in the limited time allowed for the data collection. If time had permitted, there was a room for improvement in the LDT and in the GFT.

Limitations of the lexical decision task (LDT)

The stimuli were written in capital letters. According to Baron and Strawson (1976), reading times are marginally slower for upper case than for lower case words (Coltheart, Sartori, & Job, 2013, p. 32). However, the RT in the LDT in this study was to compare the groups speed of lexical access and to compare the speed of lexical access of target words, although it would be very helpful to know the speed at which each participant would react to the target words in a more natural context.

The CG performance in the LDT might suggest the possibility of guessing in the LDT. The CG recognition rate increased in the post-test and the analysis showed that this might be due to the *testing effect*. However, the CG performance fluctuated in their recognition of target words. This might mean that they pressed on the non-word button arbitrarily, and this may have resulted in scoring for some of the target words in the LDT. This could apply also to the RMCG and the GRG, but Table 5.3 for example, shows that scores were very high for the tests (except for the CG), suggesting that the scores represents real recognition of the target words LDT.

Limitations of the Gap-fill Test (GFT)

Furthermore, the GFT had a number of limitations. Four GFTs were designed for the current study. The first GFT was used in the pilot study, and the GFT for list A was repeated in the first week, the second week, and in the delayed post-test. Similarly, the GFT for list B was

repeated in the third week and in the delayed post-test. Finally, the GFT for list C was repeated in the delayed post-test.

Each GFT included a guidance sentence of how to fill the gaps. Each gap consisted of the first two to three letters of the target word to give a clue to what needed to be written in the gaps. Any incorrect spelling was allowed up to a maximum of three spelling mistakes in one target word. It is not possible to gain information of the participant production of the target words without permitting spelling mistakes.

The study is limited in several ways, however it contributed to the research into intentional and incidental vocabulary learning and provided evidence of the effectiveness of word cards and the manual *Leitner system* and graded readers. The following are the pedagogical implications of this study.

7.5 Recommendation for Further Research

A recommendation for further research is to investigate the learners' mastery of reading skills in the L1 and compare it to the learners' mastery of reading in the L2, (e.g., guessing words from context, making inferences, predicting and comprehending the meaning of the paragraph), and the learners' ability to use their experiences and background knowledge to understand the text. After the comparison between the mastery of reading in the two languages, it would be more effective to compare two groups with different reading skills and investigate the uptake of target words incidentally.

There is a need to investigate the effect of the L1 on the uptake of target words. It is recommended to compare two groups with different L1s and to analyse learning target words intentionally or incidentally from word cards and the *Leitner system* or from graded readers, and relating this to the special difficulties which exist in the L1of the learners.

It would be beneficial to add to the computer-based word cards, especially phone applications, and to compare them to paper-based word cards used in the current study. In Nakata's (2008) study, where he compared the use of computer-based word cards with word lists and paper-based word cards, there was a significant out-performance by the computer-based word card group over the word list group, and the learners showed their preference for PC techniques over other techniques. Additionally, activities that are only suitable for computers can be created, such as activities that require fast and easy access to resources and to other learners (Nation, 2013). Furthermore, activities that require immediate feedback on success and progress, or activities that require learners to listen to the pronunciation of vocabulary items, or those that require learners to record their voices (Nation, 2013) are

suggested. In other words, the ability to have multimedia resources is beneficial to vocabulary research (Nakata, 2008; Nation, 2013).

Another characteristic of computer software is the ability to save a record of a learner's progress in learning every individual word a learner acquires, and to plan the sequence of learning, which is very useful even for learners with limited skills (N.C. Ellis, 1995; Nakata, 2008, 2011; Nation, 2013). The software can adapt to the learners' performances and apply the best conditions possible according to research findings (Nation, 2013). Applying the spaced repetition technique with computer software apparently offers a more effective spaced learning system than using word cards (Nakata, 2008). Systemizing word cards is somehow more demanding because learners have to make a number of judgments in order to divide the cards into suitable categories (i.e., easy, familiar but needs revision, difficult) (Nakata, 2008). Computer programming allows for retrieval practice and variation in the presentation of words (Nakata, 2008).

This study investigated repetition (20 minutes learning every day for three weeks except for the weekends) but much attention should be considered of other factors that can influence the validity of the repetition (Vidal, 2011). For example, learner readiness, the saliency of the form, and the pressure of learning explicitly are other factors that affect the validity of repetition (VanPatten, Williams, & Rott, 2004, p. 15).

References

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

Allen, P. A., Stadtlander, L. M., Groth, K. E., Pickle, J. L., & Madden, D. J. (2000). Adult Age Invariance in Sentence Unitization. *Aging, Neuropsychology, and Cognition*, *7*(1), 54–67.

Altmann, G. T. m. (1995). *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (New Ed edition). Cambridge, Mass.: MIT Press.

Amiri, H., Miller, T., & Savova, G. (2017). Repeat before Forgetting: Spaced Repetition for Efficient and Effective Training of Neural Networks. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2391–2400.

Anderson, J. R. (John R. (2015). *Cognitive psychology and its implications / John R. Anderson.* (8th ed.). Worth.

Anderson, R. C., & Freebody, p. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews* (pp. 77–17). Newark, DE: Interntional Reading Association.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589.

Baddeley, A. (1997). *Human Memory: Theory and Practice, Revised Edition* (2 edition). Hove, East Sussex: Psychology Press.

Baddeley, A. (1998). Working memory. *Academia Des Sciences/ Elsevier, Paris*, *321*, 167–173.

Baddeley, A. (2004). *Your Memory: A User's Guide by Baddeley, Alan D.* Prion Books Limited.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*, *105*(1), 158–173.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of Foreign Language Vocabulary and the Spacing Effect. *Psychological Science*, *4*(5), 316–321.

Barcroft, J., Schmitt, N., & Sunderman, G. (2011). Lexis. In J. Simpson (Ed.), *The Routledge Handbook of Applied Linguistics* (1 edition). Milton Park, Abingdon, UK ; New York: Routledge.

Bassetti, B., & Cook, V. (2005). An introduction to researching second language writing systems. In B. Bassetti & V. Cook (Eds.), *Second Language Writing Systems*. Multilingual Matters.

Bauer, L., & Nation, P. (1993). Word Families. *International Journal of Lexicography*, *6*(4), 253–279.

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

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society* (pp. 56–64). New York: Worth Publishers.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 396–401). London: Wiley.

Bloom, K. C., & Shuell, T. J. (1981). Effects of Massed and Distributed Practice on the Learning and Retention of Second-Language Vocabulary. *The Journal of Educational Research*, *74*(4), 245–248.

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.

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces Between Second Language Acquisition and Language Testing Research*. Cambridge University Press.

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, *41*(4), 244–262.

Carney, E. (2012). *A Survey of English Spelling*. Routledge.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction. *Educational Psychology Review*, *24*(3), 369–378.

Carter, R., & McCarthy, M. (1988). *Vocabulary and Language Teaching* (1 edition). London ; New York: Routledge.

Catford, J. C., Palmer, J. D., McCarus, E., Moray, E., & Sinder, S. A. (1974). A contrastive study of English and Arabic. University of Michigan.

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.

Claridge, G. (2005). Simplification in graded readers: Measuring the authenticity of graded texts - Linguistics and Language Behavior Abstracts (LLBA) - ProQuest. *Reading in a Foreign Language*, *17*(2), 144–158.

Cobb, T. (2007). Computing the vocabulary demands of l2 reading. *Language Learning & Technology*, *11*(3), 38–64.

Cobb, T. (2008). What the reading rate research does not show: Response to McQuillan & Krashen. *Language Learning & Technology*, *12*(1), 109–114.

Cobb, T. (2015). VOCAB SIZE TEST. Retrieved 10 November 2017, from https://www.lextutor.ca/tests/levels/recognition/1_14k/

Cobb, T., & Horst, M. (2011). Does Word Coach Coach Words? *CALICO Journal*, *28*(3), 639–661.

Cohen, A. D. (1987). The Use of Verbal and Imagery Mnemonics in Second-Language Vocabulary Learning. *Studies in Second Language Acquisition*, *9*(1), 43–61.

Coltheart, M., Sartori, G., & Job, R. (Eds.). (2013). *The Cognitive Neuropsychology of Language* (1 edition). Place of publication not identified: Psychology Press.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, *34*.

Cull, W. L., Shaughnessy, R., J. J., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied*, *2*(4).

Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and Assessing Vocabulary Knowledge*. Cambridge ; New York: Cambridge University Press.

Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, *7*(2), 541–551.

De Cock, S., & Granger, S. (2004). High frequency words: The bête noire of lexicographers and learners alike. A close look at the verb make in five monolingual learners' dictionaries of English. In G. Williams & S. Vessier (Eds.), *Proceedings of the 11th EURALEX International Congress* (pp. 233–243).

de Groot, A. M. B. (2006). Effects of Stimulus Characteristics and Background Music on Foreign Language Vocabulary Learning and Forgetting. *Language Learning*, *56*(3), 463–506.

Dekeyser, R. (2003). Implicit and Explicit Learning. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (New edition edition, pp. 313–348). Wiley-Blackwell.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*(8), 627–634.

Dempster, F. N., & Perkins, P. G. (1993). Revitalizing classroom assessment: Using tests to promote learning. *Journal Of Instructional Psychology*, *20*(3), 197.

Diack, H. (1975). *Test Your Own Word Power*. London: Paladin.

Dosher, B. (2005). Working Memory. In *Ensyclopedia of cognitive science*. Retrieved from http://search.credoreference.com.idpproxy.reading.ac.uk/content/entry/wileycs/working_memory/0

Dupuy, B., & Krashen, S. D. (1993). Incidental Vocabulary Acquisition in French as a Foreign Language. *Applied Language Learning*, *4*, 55–63.

Educational Testing Service. (1995). *TOEFL sample test 5th ed*. NJ: Princeton.

Eldridge, J., & Neufeld, S. (2009). The Graded Reader Is Dead, Long Live the Electronic Reader. *Reading Matrix: An International Online Journal*, *9*(2), 224–244.

Elgort, I. (2011). Deliberate Learning and Vocabulary Acquisition in a Second Language. *Language Learning*, *61*(2), 367–413.

Ellis, N. C. (1995). The Psychology of Foreign Language Vocabulary Acquisition: Implications for Call. *Computer Assisted Language Learning*, *8*(2–3), 103–128.

Ellis, R. (1997). *Second Language Acquisition*. (H. G. Widdowson, Ed.). London: OUP Oxford.

Engel de Abreu, P. M. J., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, *38*(6), 552–561.

English Dictionary, Thesaurus, & grammar help | Oxford Dictionaries. (n.d.). Retrieved 21 November 2017, from https://en.oxforddictionaries.com/

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*(2), 289–315.

Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language; Honolulu*, *20*(1), 19–42.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.

Field, J. (2003). *Psycholinguistics: A Resource Book for Students*. London: Routledge.

Finkenbinder, E. O. (1913). *The Curve of Forgetting*. The American Journal of Psychology.

Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive vocabulary learning: a case study. *Language Learning Journal*, *36*(2), 239–248.

Folse, K. S. (2004). *Vocabulary Myths: Applying Second Language Research to Classroom Teaching*. University of Michigan Press.

Fromkin, V., Rodman, R., & Hyams, N. (2010). *An Introduction to Language, 9th Edition* (9th edition). Boston, MA: Wadsworth Cengage Learning.

Gleitman, L. R., & Gleitman, H. (1994). 'A picture is worth a thousand words, but that's the problem'. In B. Lust, M. Suner, & J. Whitman (Eds.), *The Role of Syntax in Vocabulary Acquisition: Crosslinguistic Perspectives.* Hove, UK: LEA Publishers.

Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: Approaches to second-language vocabulary learning and technology. *Language Learning and Technology*, *14*(2), 4–11.

Godwin-Jones, R. (2011). Emerging Technologies Mobile Apps for Language Learning - Semantic Scholar. *Language Learning & Technology*, *15*(2), 2–11.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., Bouwmeester, S., & Zwaan, R. A. (2016). Distributed Practice and Retrieval Practice in Primary School

Vocabulary Learning: A Multi-classroom Study. *Applied Cognitive Psychology*, *30*(5), 700–712.

Goulden, R., Nation, P., & Read, J. (1990). How Large Can a Receptive Vocabulary Be? *Applied Linguistics*, *11*(4), 341–363.

Hale, G. A., Stansfield, C. W., Rock, D. A., Hicks, M. M., Butler, F. A., & Oller, J. W. (1989). The relation of multiple-choice cloze items to the Test of English as a Foreign Language. *Language Testing*, *6*(1), 47–76.

Harley, T. A. (2008). *The Psychology of Language from data to theory*. NY: Psychology Press.

Helland, T., & Asbj⊘rnsen, A. (2004). Digit Span in Dyslexia: Variations According to Language Comprehension and Mathematics Skills. *Journal of Clinical and Experimental Neuropsychology*, *26*(1), 31–42.

Hill, M., & Laufer, B. (2006). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL - International Review of Applied Linguistics in Language Teaching*, *41*(2), 87–106.

Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, *8*(2), 689–696.

Horn, J. L. (1982). The Theory of Fluid and Crystallized Intelligence in Relation to Concepts of Cognitive Psychology and Aging in Adulthood. In F. Craik (Ed.), *Aging and Cognitive Processes* (pp. 237–278). Springer, Boston, MA.

Hornung, C., Brunner, M., Reuter, R. A. P., & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence*, *39*(4), 210–221.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a Foreign Language*, *11*(2), 207–23.

Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive on-line database. *Language Learning and Technology*, *9*(2), 90–110.

How Many Syllables. (n.d.). Retrieved 21 November 2017, from https://www.howmanysyllables.com

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258–286). Cambridge: Cambridge University Press.

Hulstijn, J. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (New edition edition, pp. 318–349). Wiley-Blackwell.

Hulstijn, J. (2010). measuring second language proficiency. In E. Blom & S. Unsworth (Eds.), *Experimental Methods in Language Acquisition Research* (UK ed. edition, pp. 185–199). Philadelphia: John Benjamins Publishing Company.

Hung, H.-T. (2015). Intentional Vocabulary Learning Using Digital Flashcards. *English Language Teaching*, *8*(10), 107–112.

Hunt, R. R., & Ellis, H. (2003). *Fundamentals of Cognitive Psychology* (7 edition). Boston: McGraw-Hill Humanities/Social Sciences/Languages.

Jacobs, H. L., Zingraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL Composition: A Practical Approach. English Composition Program*. Newbury House Publishers, Inc.

Jayaraman, S. (2010). Acquisition of pronunciation of consonant clusters by Arabic speakers of English as a second language. *Sino-US English Teaching*, *7*(1), 46–56.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning Vocabulary Through Reading. *American Educational Research Journal*, *21*(4), 767–787.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to

verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General*, *133*(2), 189–217.

Kang, S. (2016). Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. *SAGE Journals*, *3*(1), 12–19.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*(9), 1297–1317.

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. D. (2003). *Explorations in Language Acquisition and Use*. Portsmouth, N.H: Heinemann.

Kremmel, B., & Schmitt, N. (2016). Interpreting Vocabulary Test Scores: What Do Various Item Formats Tell Us About Learners' Ability to Employ Words? *Language Assessment Quarterly*, *13*(4), 377–392.

Lado, R. (1964). *Language teaching, a scientific approach.* New York: McGraw-Hill.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Laufer, B. (1988). The concept of 'synforms' (similar lexical forms) in vocabulary acquisition. *Language and Education*, *2*(2), 113–132.

Laufer, B. (1988). What percentage of text-lexis is essntial for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: From Human Thinking to Thinking Machines*. Clevedon England ; Philadelphia: Multilingual Matters Ltd.

Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*, *6*, 10–20.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Laurén & M. Nordman (Eds.), *Special Language: From Humans Thinking to Thinking Machines* (pp. 316–323). Clevedon: Multilingual Matters.

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, *16*(3), 307–322.

Laufer, B. (1996). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In *Second Language Vocabulary Acquisition*. Cambridge University Press.

Laufer, B. (2003). Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *59*(4), 567–587.

Laufer, B. (2005). Focus on Form in Second Language Vocabulary Learning. *EUROSLA Yearbook*, *5*, 223–250.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30.

Laufer, B., & Sim, D. D. (1985). Taking the easy way out: non-use and misuse of clues in EFL reading. *English Teaching Forum*, *23*(2), 7–20.

Laws, J., & Ryder, C. (2014). Getting the Measure of Derivational Morphology in Adult Speech: A Corpus Analysis Using MorphoQuantics. *Language Studies Working Papers*, *6*, 3–17.

Lewandowsky, S. (2005). Working memory, computational models of. In *Encyclopedia of cognitive science*. Retrieved from http://search.credoreference.com.idpproxy.reading.ac.uk/content/entry/wileycs/working_memory_computational_models_of/0

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Hove, UK: Language Teaching Publications.

Liu, N., & Nation, I. S. P. (1985). Factors Affecting Guessing Vocabulary in Context. *RELC Journal*, *16*(1), 33–42.

Logan, J. M., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning*, *7*(3), 175–195.

Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, *25*(1), 91–102.

Meara, P. (1990). A note on passive vocabulary. *Interlanguage Studies Bulletin (Utrecht)*, *6*(2), 150–154.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and Competence in Second Language Acquisition*. Cambridge ; New York: Cambridge University Press.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606.

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol, UK ; Buffalo N.Y.: Multilingual Matters.

Milton, J. (2016). How Different Is Arabic from Other Languages? *Journal of Applied Linguistics and Language Research*, *3*(1), 15.

Mondria, J.-A. (2007). Myths about vocabulary acquisition. *Babylonia: The Swiss Journal of Language Teaching and Learning*, *2*, 63–68.

Mondria, J.-A., & Mondria-De Vries, S. (1994). Efficiently memorizing words with the help of word cards and 'hand computer': Theory and applications. *System*, *22*(1), 47–57.

Morris, L., & Cobb, T. (2004). Vocabulary Profiles as Predictors of the Academic Performance of Teaching English as a Second Language Trainees. *System: An International Journal of Educational Technology and Applied Linguistics*, *32*(1), 75–87.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, *20*(2), 233–253.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: implications from cognitive psychology research for optimal spaced learning. *ReCALL*, *20*(1), 3–20.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: a critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*(1), 17–38.

Nation, I. S.P. (2006). How Large a Vocabulary is Needed For Reading and Listening? *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–82.

Nation, I. S. P. (1980). Strategies for receptive vocabulary learning. *Guidelines*, *3*, 18–23.

Nation, I. S. P. (1990). *Teaching & Learning Vocabulary*. Boston, Mass: Heinle Thomson ELT.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge ; New York: Cambridge University Press ELT.

Nation, I. S. P. (2002). Best practice in vocabulary teaching and learning. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in Language Teaching: An Anthology of Current Practice* (pp. 267–272). New York: Cambridge University Press.

Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, *44*(4), 529–539.

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

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*. Cambridge ; New York: Cambridge University Press.

Nation, I. S. P. (2014a). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*(2), 1–16.

Nation, I. S. P. (2014b). Prof Paul Nation | School of Linguistics and Applied Language Studies | Victoria University of Wellington. Retrieved 4th of September 2017, from https://www.victoria.ac.nz/lals/about/staff/paul-nation#free-graded-readers

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Journal Teacher*, *31*(7), 9–13.

Nation, I. S. P., & Wang Mingtzu, K. (1999). Graded Readers and Vocabulary. *Reading in a Foreign Language*, *12*(2), 355–380.

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Nation, I. S. P., & Coady, J. (1988). Vocabulary and reading. In R. Carter & M. Mccarthy (Eds.), *Vocabulary and Language Teaching* (1 edition, pp. 97–110). London ; New York: Routledge.

Nation, I. S. P., & Webb, S. (2017). *How Vocabulary Is Learned*. Oxford, New York: Oxford University Press.

Nunan, P. D. (2012). *What Is This Thing Called Language?* (2nd edition edition). New York: Palgrave Macmillan.

Oxford, R. (1990). *Language Learning Strategies: What Every Teacher Should Know*. Boston, Mass: Heinle ELT.

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.

Pimsleur, P. (1967). A Memory Schedule. *The Modern Language Journal*, *51*(2), 73–75.

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.

Prince, P. (1996). Second Language Vocabulary Learning: The Role of Context versus Translations as a Function of Proficiency. *The Modern Language Journal*, *80*(4), 478–493.

Randall, M., & Meara, P. (1988). How Arabs Read Roman Letters. *Reading in a Foreign Language*, *4*(2), 133–45.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning*, 11–18.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press

Read, J. (2007). Second Language Vocabulary Assessment: Current Practices and New Directions. *International Journal of English Studies*, *7*(2), 105−126.

Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, *18*(1), 1−32.

Reddy, S., Labutov, I., Banerjee, S., & Joachims, T. (2016). Unbounded Human Learning: Optimal Scheduling for Spaced Repetition.

Richards, J. C., & Schmidt, R. W. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. Routledge.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention - ScienceDirect. *Trends in Cognitive Sciences*, *15*(1), 20–27.

Roediger, H. L., & Karpicke, J. D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*(3), 181–210.

Rogers, J. (2017). The Spacing Effect and its Relevance to Second Language Acquisition. *Applied Linguistics*.

Rohrer, D., & Pashler, H. (2007). Increasing Retention Without Increasing Study Time. *Current Directions in Psychological Science*, *16*(4), 183–186.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463.

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics; New York*, *24*(3), 431–451.

Saigh, K., & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System*, *40*(1), 24–36. https://doi.org/10.1016/j.system.2012.01.005

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*(2), 72–78.

Scarborough, H. S., & Brady, S. A. (2002). Toward a Common Terminology for Talking about Speech and Reading: A Glossary of the 'Phon' Words and Some Related Terms. *Journal of Literacy Research*, *34*(3), 299–336.

Schacter, D. L. (2002). *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Boston: Houghton Mifflin.

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning : a behavioral emphasis / Richard A. Schmidt, Timothy D. Lee.* (4th edition.). Human Kinetics.

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: theoretical underpinnings and practical suggestions. *ELT Journal*, *49*(2), 133–143.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge ; New York: Cambridge University Press.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Springer.

Schmitt, N. (Ed.). (2010). In *Researching Vocabulary: A Vocabulary Research Manual*. Houndmills, Basingstoke, Hampshire; New York, NY: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: what the research shows. *Language Learning*, *64*(4), 913–951.

Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory. *Language Teaching Research*, *19*(1), 28–42.

Seashore, R. H., & Eckerson, L. D. (1940). The measurement of individual differences in general English vocabularies. *Journal of Educational Psychology*, *31*(1), 14–38.

Shefelbine, J. L. (1990). Student Factors Related to Variability in Learning Word Meanings from Context. *Journal of Reading Behavior*, *22*(1), 71–97.

Sokmen. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 237–257). Cambridge; New York: Cambridge University Press.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: when and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*(2), 244–253.

Swan, M., & Smith, Bernard (Eds.). (2001). *Learner English: A Teacher's Guide to Interference and other Problems* (2 edition). Cambridge; New York: Cambridge University Press.

Thomas, M. H., & Dieter, J. N. (1987). The Positive Effects of Writing Practice on Integration of Foreign Words in Memory. *Journal of Educational Psychology*, *79*(3), 249–53.

Traxler, M. J. (2011). *Introduction to Psycholinguistics: Understanding Language Science* (1 edition). Chichester, West Sussex ; Malden MA: Wiley-Blackwell.

Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: how measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*

Tulving, E., & Craik, F. (Eds.). (2000). *The Oxford Handbook of Memory*. Oxford ; New York: Oxford University Press.

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic Neighborhood Effects in Bilingual Word Recognition. *Journal of Memory and Language*, *39*(3), 458–483.

VanPatten, B., Williams, J., & Rott, S. (2004). Form-Meaning Connections in Second Language Acquisition. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-Meaning Connections in Second Language Acquisition* (pp. 1–26). Mahwah, NJ: Erlbaum.

Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards & B. Laufer-Dvorkin (Eds.), *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (pp. 173–189). John Benjamins Publishing.

Vidal, K. (2011). A Comparison of the Effects of Reading and Listening on Incidental Vocabulary Acquisition. *Language Learning*, *61*(1), 219–258.

Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: Sufficient but not necessary. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 94–105.

Waring, R. (1997). Graded and Extensive Reading: Questions and Answers. *LANGUAGE TEACHER-KYOTO-JALT-*, *21*, 9–12.

Waring, R. (2004). In defence of learning words in word pairs but only when doing it the 'right' way! Retrieved from http://www.robwaring.org/vocab/principles/systematic_learning.htm

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*(2), 130–163.

Warren, P. (2012). *Introducing Psycholinguistics*. Cambridge University Press.

Webb, S. (2007a). Learning word pairs and glossed sentences: the effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63–81.

Webb, S. (2007b). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics*, *28*(1), 46–65.

Westphal, P. B. (1977). In Search of a Systematic Way with Words. *The French Review*, *51*(1), 59–64.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6–10.

Yang, L. (1997). Tracking the acquisition of L2 vocabulary: the Keki language experiment. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge University Press.

Appendices

Appendix 1. Research Documents
A. Ethics Documents/ Project Submission

School of Literature and Languages
Department of English Language and Applied Linguistics

University of Reading

ETHICS COMMITTEE

Principal Investigator (Supervisor): Prof. Jeanine Treffers-Daller

Student name: Sally Alghamdi

Department: Department of English Language and Applied Linguistics

Title of Project: The effect of using spaced repetition technique on vocabulary retention in English for second language learners.

Proposed starting date: 1/9/2015

Number of participants that you require consent from (approximate): 150 participants.

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…….

Sally Alghamdi(Student) Date 19 May 2015

Appendix 1. Research Documents
A. Ethics Documents/ Consent Form

School of Literature and Languages
Department of English Language and Applied Linguistics

University of Reading

Project title: The effect of using spaced repetition technique on vocabulary retention in English for second language learners.

I understand the purpose of this research and understand what is required of me; I have read and understood the Information Sheet relating to this project, which has been explained to me by Mrs.Sally Alghamdi. 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 at any time.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix 1. Research Documents
A. Ethics Documents/ Information sheet

Researcher:
Sally Alghamdi
Email: s.a.s.alghamdi@pgr.reading.ac.uk

Supervisor:
Dr Jeanine Treffers-Daller
Phone: 00-44-118-3782690
Email: j.c.treffers-daller@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
+44 (0)118 378 6472 +44 (0)118 975 6506
Email
appling@reading.ac.uk p.a.thompson@reading.ac.uk

This study is about an experiment that investigates
spacing and learning. Second language learners like you is invited to participate in the study. Also as a student in Albaha University your cooperation is important to the research sponsored by Albaha University. You are going to participate by (using word cards, reading) as instructed for a half an hour everyday this month.
The information you give will remain confidential. You will not be identifiable in any published report resulting from the study. Information about individuals will not be shared with anyone else. The data will be held in a password-protected computer and any paper work in a locked cupboard. The materials will be destroyed immediately after the completion of the researcher dissertation. You have the right to know about the results of this study. You can withdraw by contacting the researcher by email

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

Signed

Appendix 1. Research Documents
B. Department of Foreign Languages study plan

Albaha University
Faculty of Arts &
Human Sciences

Dept. of Foreign Languages
English Language Division

STUDY PLAN: B.A. IN ENGLISH LANGUAGE & LITERATURE

Level 3				Level 4		
Course	**Title**	**Hrs**		**Course**	**Title**	**Hrs**
15021201	Listening & Speaking (1)	3		15021202	Listening & Speaking (2)	3
15021203	Reading (1)	2		15021204	Reading (2)	2
15021205	Writing (1)	2		15021206	Writing (2)	2
15021207	Grammar (1)	3		15021208	Grammar (2)	3
15021209	Vocabulary Building	2		15021210	Phonetics (1)	2
15021211	Introduction to Linguistics	2		15021212	Introduction to Literature	2
15041213	السيرة النبوية	2		15021214	Translation (1)	2
15011215	فـن القراءة	2		15011216	فـن الكتابة	2
Total		18		Total		18
Level 5				**Level 6**		
Course	**Title**	**Hrs**		**Course**	**Title**	**Hrs**
15021301	Appreciating Poetry	3		15021302	Romantic Poetry	3
15021303	Appreciating Drama	2		15021304	19th Century Novel	3
15021305	Essay Writing	2		15021306	Language and Computer	2
15021307	Phonetics (2)	2		15021308	Translation (3)	2
15021309	Translation (2)	2		15021310	Phonology (1)	2
15021311	Syntax	2		15021312	Morphology	2
15021413	Research Methods	2		15021314	Semantics	2
15021315	Language Acquisition	2		15021316	16th Century Drama	3
15021317	The Rise of the Novel	2		Total		19
Total		19				
Level 7				**Level 8**		
Course	**Title**	**Hrs**		**Course**	**Title**	**Hrs**
15021403	Phonology (2)	2		15021404	Pragmatics	2
15021405	Intro. to American Literature	2		15021406	Modern Poetry	2
15021407	History of English Lang.	2		15021408	Modern Drama	2
15021409	Discourse Analysis	3		15021410	Modern British Novel	3
15021411	Victorian Poetry	3		15021412	Literary Criticism (2)	3
15021313	Literary Criticism (1)	2		15021414	Comparative Literature	2
15021415	Sociolinguistics	2		15021402	Transformational Grammar	2
15021401	Applied Linguistics	2		Total		16
Total		18				

Appendix 2. Pre-tests
A. Vocabulary Size Test

Circle the letter a-d with the closest meaning to the key word in the question.

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 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
10. BASIS: This was used as the basis.
    a. answer
    b. place to take a rest
    c. next step
    d. main part

Second 1000

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
    b. important secrets
    c. a stupid person
    d. someone who writes for a newspaper
    e. someone who is paid for playing sport etc

Third 1000

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

Fourth 1000

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. 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

Fifth 1000

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. a. break
   b. small piece
   c. short coat

d. rare jewel

10. BACTERUM: 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. a thing that has been stolen and sold

Sixth 1000

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 to a shop

Seventh 1000

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

Eighth 1000

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
9. MUMBLE: He started to mumble.

a. think deeply
b. shake uncontrollably
c. stay further behind the others
d. speak in an unclear way

Ninth 1000

1. HALLMARK: Does it have a hallmark?
   a. stamp to show when to use it by
   b. stamp to show the quality
   c. mark to show it is approved by the royal family
   d. Mark or stain to prevent copying
2. PURITAN: He is a puritan.
   a. person who likes attention
   b. person with strict morals
   c. person with a moving home
   d. person who hates spending money
3. MONOLOGUE: Now he has a monologue.
   a. single piece of glass to hold over his eye to help him to see better
   b. long turn at talking without being interrupted
   c. position with all the power
   d. picture made by joining letters together in interesting ways
4. WEIR: We looked at the weir.
   a. person who behaves strangely
   b. wet, muddy place with water plants
   c. old metal musical instrument played by blowing
   d. thing built across a river to control the water
5. WHIM: He had lots of whims.
   a. old gold coins
   b. female horses
   c. strange ideas with no motive
   d. sore red lumps
6. PERTURB: I was perturbed.
   a. made to agree
   b. worried
   c. very puzzled
   d. very wet
7. REGENT: They chose a regent.
   a. an irresponsible person
   b. a person to run a meeting for a time
   c. ruler acting in place of the king
   d. a person to represent them
8. OCTOPUS: They saw an octopus.
   a. a large bird that hunts at night
   b. a ship that can go under water
   c. a machine that flies by means of turning blades
   d. a sea creature with eight legs
9. FEN: The story is set in the fens.
   a. low land partly covered by water
   b. a piece of high land with few trees
   c. a block of poor-quality houses in a city
   d. a time long ago
10. LINTEL: He painted the lintel.
    a. beam over the top of a door or window
    b. small boat used for getting to land from a big boat
    c. beautiful tree with spreading branches and green fruit
    d. board showing the scene in a theatre

Tenth 1000

1. AWE: They looked at the mountain with awe.
   a. worry
   b. interest
   c. wonder
   d. respect
2. PEASANTRY: He did a lot for the peasantry.
   a. local people
   b. place of worship
   c. businessmen's club
   d. poor farmers
3. EGALITARIAN: This organization is egalitarian.
   a. does not provide much information about itself to the public
   b. dislikes change
   c. frequently asks a court of law for a judgement
   d. treats everyone who works for it as if they are equal
4. MYSTIQUE: He has lost his mystique.
   a. his healthy body

b. the secret way he makes other people think he has special power or skill
c. the woman who has been his lover while he is married to someone else
d. the hair on his top lip

5. UPBEAT: I'm feeling really upbeat about it.
   a. upset
   b. good
   c. hurt
   d. confused

6. CRANNY: We found it in the cranny!
   a. sale of unwanted objects
   b. narrow opening
   c. space for storing things under the roof of a house
   d. large wooden box
7. PIGTAIL: Does she have a pigtail?
   a. a rope of hair made by twisting bits together
   b. a lot of cloth hanging behind a dress
   c. a plant with pale pink flowers that hang down in short bunches
   d. a lover
8. CROWBAR: He used a crowbar.
   a. heavy iron pole with a curved end
   b. false name
   c. sharp tool for making holes in leather
   d. light metal walking stick
9. RUCK: He got hurt in the ruck.
   a. hollow between the stomach and the top of the leg
   b. pushing and shoving
   c. group of players gathered round the ball in some ball games
   d. race across a field of snow
10. LECTERN: He stood at the lectern.
    a. desk to hold a book at a height for reading
    b. table or block used for church sacrifices
    c. place where you buy drinks
    d. very edge

Eleventh 1000

1. EXCRETE: This was excreted recently.
   a. pushed or sent out
   b. made clear
   c. discovered by a science experiment
   d. put on a list of illegal things
2. MUSSEL: They bought mussels.
   a. small glass balls for playing a game shellfish
   b. large purple fruits
   c. pieces of soft paper to keep the
   d. clothes clean when eating
3. YOGA: She has started yoga.
   a. handwork done by knotting thread
   b. a form of exercise for body and mind
   c. a game where a cork stuck with feathers is hit between two players
   d. a type of dance from eastern countries
4. COUNTERCLAIM: They made a counterclaim.
   a. a demand made by one side in a law case to match the other side's demand
   b. a request for a shop to take back things with faults
   c. An agreement between two companies exchange work
   d. a top cover for a bed
5. PUMA: They saw a puma.
   a. small house made of mud bricks b. tree from hot, dry countries
   b. very strong wind that sucks up
   c. anything in its path
   d. large wild cat
6. PALLOR: His pallor caused them concern.
   a. his unusually high temperature
   b. his lack of interest in anything
   c. his group of friends
   d. the paleness of his skin
7. APERITIF: She had an aperitif.
   a. a long chair for lying on with just one place to rest an arm
   b. a private singing teacher
   c. a large hat with tall feathers
   d. a drink taken before a meal
8. HUTCH: Please clean the hutch.

a. thing with metal bars to keep dirt out of water pipes
b. space in the back of a car for bags
c. metal piece in the middle of a bicycle wheel
d. cage for small animals

9. EMIR: We saw the emir.
   a. bird with long curved tail feathers
   b. woman who cares for other people's
   c. children in Eastern countries
   d. Middle Eastern chief with power in his d. house made from blocks of ice
10. HESSIAN: She bought some hessian.
   a. oily pinkish fish
   b. stuff producing a happy state of mind c. coarse cloth
   c. strong-tasting root for flavouring food
   d. land
   e.

Twelfth 1000

1. HAZE: We looked through the haze.
   a. small round window in a ship
   b. unclear air
   c. strips of wood or plastic to cover a window
   d. list of names
2. SPLEEN: His spleen was damaged.
   a. knee bone
   b. organ found near the stomach
   c. pipe taking waste water from a house
   d. respect for himself
3. SOLILOQUY: That was an excellent soliloquy!
   a. song for six people
   b. short clever saying with a deep meaning
   c. entertainment using lights and music
   d. speech in the theatre by a character who is alone
4. REPTILE: She looked at the reptile.
   a. old hand-written book
   b. animal with cold blood and a hard-outside person who sells things by knocking on
   c. doors
   d. picture made by sticking many small pieces of different colours together
5. ALUM: This contains alum.
   a. a poisonous substance from a common plant
   b. a soft material made of artificial threads
   c. a tobacco powder once put in the nose
   d. a chemical compound usually involving aluminium
6. REFECTORY: We met in the refectory.
   a. room for eating
   b. office where legal papers can be signed
   c. room for several people to sleep in
   d. room with glass walls for growing plants
7. CAFEINE: This contains a lot of caffeine.
   a. a substance that makes you sleepy
   b. threads from very tough leaves
   c. ideas that are not correct
   d. a substance that makes you excited
8. IMPALE: He nearly got impaled.
   a. charged with a serious offence
   b. put in prison
   c. stuck through with a sharp instrument
   d. involved in a dispute
9. COVEN: She is the leader of a coven.
   a. a small singing group
   b. a business that is owned by the workers
   c. a secret society
   d. a group of church women who follow a strict religious life
10. TRILL: He practised the trill.

a. ornament in a piece of music
b. type of stringed instrument
c. Way of throwing a ball
d. dance step of turning round very fast on the toes

Thirteenth 1000

1. UBIQUITOUS: Many weeds are ubiquitous.
   a. are difficult to get rid of
   b. have long, strong roots
   c. are found in most countries
   d. die away in the winter
2. TALON: Just look at those talons!
   a. high points of mountains

b. sharp hooks on the feet of a hunting bird
c. heavy metal coats to protect against weapons
d. people who make fools of themselves without realizing it

3. ROUBLE: He had a lot of roubles.
   a. very precious red stones
   b. distant members of his family
   c. Russian money
   d. moral or other difficulties in the mind
4. JOVIAL: He was very jovial.
a. low on the social scale
b. likely to criticize others
c. full of fun
d. friendly

5. COMMUNIQUE: I saw their communiqué .
   a. a critical report about an organization
   b. garden owned by many members of a community
   c. printed material used for advertising
   d. official announcement
6. PLANKTON: We saw a lot of plankton.
   a. poisonous weeds that spread very quickly
   b. very small plants or animals found in water
   c. trees producing hard wood
   d. grey clay that often causes land to slip
7. SKYLARK: We watched a skylark.
   a. show with aeroplanes flying in patterns
   b. man-made object going round the earth
   c. person who does funny tricks
   d. small bird that flies high as it sings
8. BEAGLE: He owns two beagles.
   a. fast cars with roofs that fold down b. large guns that can shoot many
   b. people quickly
   c. small dogs with long ears
   d. houses built at holiday places
9. ATOLL: The atoll was beautiful.
   a. low island made of coral round a sea-water lake
   b. work of art created by weaving pictures from fine thread
   c. small crown with many precious jewels worn in the evening by women
   d. place where a river flows through a narrow place full of large rocks
10. DIDACTIC: The story is very didactic.
    a. tries hard to teach something
    b. is very difficult to believe
    c. deals with exciting actions
    d. is written in a way which makes the reader unsure of the meaning

Fourteenth 1000

1. CANONICAL: These are canonical examples.
   a. examples which break the usual
   b. rules examples taken from a
   c. religious book regular and widely accepted examples
   d. examples discovered very recently
2. ATOP: He was atop the hill.
   a. at the bottom of
   b. at the top of
   c. on this side of
   d. on the far side of
3. MARSUPIAL: It is a marsupial.
   a. an animal with hard feet
   b. a plant that grows for several years
   c. a plant with flowers that turn to face the sun
   d. an animal with a pocket for babies
4. AUGUR: It augured well.
   a. promised good things for the future
   b. agreed well with what was expected
   c. had a colour that looked good with something else
   d. rang with a clear, beautiful sound
5. BAWDY: It was very bawdy.
   a. unpredictable
   b. enjoyable
   c. rushed
   d. rude

6. GAUCHE: He was gauche.
    a. talkative
    b. flexible
    c. awkward
    d. determined
7. THESAURUS: She used a thesaurus.
    a. a kind of dictionary
    b. a chemical compound
    c. a special way of speaking
    d. an injection just under the skin
8. ERYTHROCYTE: It is an erythrocyte.
    a. a medicine to reduce pain
    b. a red part of the blood
    c. a reddish white metal
    d. a member of the whale family
9. CORDILLERA: They were stopped by the cordillera.
    a. a special law
    b. an armed ship
    c. a line of mountains
    d. the eldest son of the king
10. LIMPID: He looked into her limpid eyes.
    a. a clear
    b. tearful
    c. deep brown
    d. beautiful

Appendix 2. Pre-tests
B. Grammar Test

Answer all the questions (60 points)

I. Circle the correct answer (5 points)

1. Some of the fruit in the ball **(is, are)** rotten.
2. Half of this money **(is, are**) yours.
3. Each boy in the classroom (**has, have**) his own notebook.
4. The professor and the student (**agree, agrees**) on this point.
5. There (**is , are**) a lot of problems in the world.

II. Write the Plural of the following nouns. (10points)

	singular	Plural
1	Man	
2	Life	
3	Memo	
4	Deer	
5	Loaf	
6	Radio	
7	Thesis	
8	Mouse	
9	Furniture	
10	chair	

III. Fill in the blank with the correct from of the noun in the list. (10 points)

Traffic	Phenomenon	Match
River	Shoe	Truck
equipment	computer	baby

1. **I need some _____ to light the fire.**
2. **Thunder and lightning are two _____ of nature**
3. **The Mississippi, the Amazon, and the Nile are well-know _____**
4. **The street is full of cars, _____ and buses. In other words it is full of _____**
5. **They sell _____ at that store. It is a shoe store.**
6. **People can buy special food in small jars for _____. It is called _____ food. (baby)**
7. **There are _____, copiers, telephones, and staplers in a typical office. A business office needs a lot of _____.**

IV. Complete the sentences with either Since, or For. (5 points)

1. They have known each other _____more than fifty years.
2. I've had a cold _____over a week.
3. We haven't seen Aziz _____last month.
4. There has been snow on the ground _____New Year's Day.
5. Maria has studied English _____less than a year.

V. Fill in the blank with the correct article. (the, a, an or Zero)(10 points

1. We need to get _____ new phone
2. Listening to _____ loud rock music is fun.

3. _____vocabulary in this exercise is easy.
4. A: you'd better have _____good reason for being late. B: I do.
5. Lucy is wearing _____ straw hat today.
6. _____Island is a piece of land surrounded by water.
7. _____sentence usually contains a subject and a verb.
8. _____ basketball is round.
9. _____ gold is Metal.
10. Helen got _____ letter from her mother.

VI. Write the correct form of the verbs in parentheses. (5 point)

1. Tom (watch) _____TV every day.
2. I (buy) _____ a textbook yesterday.
3. Tom (sleep) _____ right now.
4. Ahmad (go) _____ to school yesterday.
5. Tom had some work to do, so he (study) _____ last night.

VII. Select the most appropriate answer. (15 points)

1. According to the weather report it (be) _____ cloudy tomorrow.
(a) Will be (b) are raining (c) has raining (d) will rain
2. I was late. The teacher (give, already)_____ a quiz when I got to class.
(a) Has already given (b) will be given (c) had already given (d) gave
3. It (rain) _____ all day.
(a) Has been raining (b) had raining (c) had been raining (d) had rain
4. By the time she got there, Ann (left) _____
(a) Had left (b) will leave (c) will have left (d) is leaving
5. Last January, I (see) _____snow for the first time in my life.
(a) Will see (b) had seen (c) saw (d) has seeing
6. Mary always (leave) _____ for school at 7:45.
(a) Left (b) will leave (c) leaves (d) leaving
7. I (come) _____ tomorrow around five o'clock.
(a) Am coming (b) was come (c) will come (d) has come
8. I (buy) _____a new car last week.
(a) Was buy (b) are buying (c) bought (d) had buyed
9. It was raining hard, but by the time class (be) _____ over, the rain had stopped.
(a) Is (b) was (c) has been (d) are being
10. While I (walk) _____ down the street, it began to rain.
(a) was walking (b) is walking (c) walking (d) had walk
11. Eric finally came home at six O'clock. I (wait) _____ for him since four thirty.
(a) Have waited (b) is wait (c) had been waiting (d) had wait
12. We (have) _____ three tests so far this term.
(a) Are taking (b) have took (c) have had (d) taking
13. Samir was a newspaper reporter before he (become) _____ a businessman.
(a) Become (b) is become (c) became (d) has become
14. I (know) _____ Bob since I was in high school.

(a) Have known (b) was knowing (c) knew (d) had know

15. I will get home around six. When I get home I (give) _____ Karen a call.

(a) Will give (b) gave (c) has given (d) had given

Appendix 2. Pre-tests
C. Forward Digit Span Test

English

Task 1: I am going to say some numbers. Listen carefully, and when I am done, say them after me.

6 - 1 – 2

7 – 9 – 5

8 – 6 - 7

6 – 1 - 5 – 8

9 – 0 – 8 – 1

7 – 6 – 3 – 4

5 - 2 - 1 - 8 – 6

4 – 7 – 5 – 9- 8

1 -5 – 6 – 3 - 4

7 - 9 - 6 - 4 - 8 – 3

6 – 5 – 2 – 9 – 6 -7

5 – 4 – 3 – 8 - 3 - 5

9 - 8 - 5 - 2 - 1 - 6 – 3

6- 5 – 3 – 9 – 4 – 7- 1

2 – 4 – 6 – 8 -5 – 8 - 9

2 - 9 - 7 - 6 - 3 - 1 - 5 – 4

8 – 6 – 4 – 5 – 8- 9 – 7 -3

5 – 4 – 6 – 2 – 7 – 3 – 9 – 1

Arabic

Task 2: I am going to say some numbers. Listen carefully, and when I am done, say them after me.

١ـ ٣ـ ٥

٨ ـ٦ـ ٩

٢ـ ٥ـ٧

٣ ـ٨ـ ٧ـ ٩

٤ ـ١ـ ٣ـ ٦

٢ ـ٩ـ ٦ـ ٥

٢ ـ٣ـ ٥ـ ٧ـ ٨

٣ ـ٢ـ ٧ـ ٦ـ ١

٦ ـ٨ـ ٣ـ ٩ـ ٦

٤ ـ٩ـ ٨ـ ٢ـ ٥ـ ٧

٧ ـ٩ـ ٣ـ ٩ـ ٦ـ ٤

٤ ـ٥ـ ٨ـ ٩ـ ٧ـ ٣

٧ ـ١ـ ٣ـ ٨ـ ٦ـ ٤ـ ٩

٤ ـ١ـ ٩ـ ٨ـ ٧ـ ٥ـ ٦

١ ـ٢ـ ٦ـ ٣ـ ٩ـ ٨ـ ٤

٢ـ٩ـ ٧ـ ٣ـ ٤ـ ٦ـ ٩ـ ٥

٣ ـ٦ـ ٢ـ ٥ـ ٧ـ ٩ـ ٤ـ ١

٨ ـ١ـ ٣ـ ٩ـ ٧ـ ٥ـ ٤ـ ٦

Appendix 2. Pre-tests
D. Questionnaire

This study investigates the potential effects of a spaced repetition technique on the retention of information, by comparing results from four groups of learners at Albaha University. The comparison explores the potential effects of rehearsal as well as that of planned repetition on the L2 learner memories by using different techniques.

"By completing and returning this questionnaire I understand that I am giving consent for my responses to be used for the purposes of this research project".

Name:……..

Signature:……

Name: _____

Part 1: Background Information

1. Gender

    a. Male
    b. Female

2. Age: _____

3. Country of birth: _____

4. What is your native language?

    a. Arabic
    b. English
    c. Other

5. What languages do you speak at home?

    a. Arabic
    b. English
    c. Other

6. What year are you in the school of English Language and Literature?

    a. Year 1
    b. Year 2
    c. Year 3
    d. Year 4
    e. Year5

7. In what languages did you receive the majority of your precollege education?

    a. Arabic
    b. English
    c. Other

8. Have you ever been to an English-speaking region for the purpose of studying English?

    a. Yes
    b. No

8.1 If yes, when?

_____

8.2 Where?

_____

8.3 For how long?

    a. semester or less
    b. 2 semesters
    c. more than 2 semesters

9. Other than the experience mentioned in Question 7, have you ever lived in a situation where you were exposed to English E.g. by living in an English speaking community com; visiting English community for purposes of study abroad or work; exposure through family members, etc. ?

    a. Yes
    b. No

If Yes, please give details below.

9.1 Country/ region

_____
_____

9.2 Purpose

_____
_____
_____
_____
_____
_____

9.3 From when to when

_____
_____

10. Rate your language ability in Arabic

a. Poor
b. Good
c. Very good
d. Native/native-like

11. How many years (if any) have you studied Arabic in a formal school setting?

_____

12. Rate your language ability in English

a. Poor
b. Good
c. Very good
d. Native/native-like

13. How many years (if any) have you studied English in a formal school setting?

_____

14. Have you studied English in school in the past at each of the levels listed below? how long?

14.1 Elementary school:

_ No _Yes

a. less than 1 year
b. 1–2 years
c. more than 2 years

14.2 Junior high (middle) school:

_No _Yes

a. less than 1 year
b. 1–2 years
c. more than 2 years

14.3 Senior high school:

_No _Yes

a. less than 1 year
b. 1–2 years
c. more than 2 years

14.4 University college (other than English and Literture)

_No _Yes

a. less than 1 year
b. 1–2 years
c. more than 2 year

Part 2: All of the Questions That Follow Refer to Your Use of <u>English</u> only

Rate the following statements:

1. On average I communicate with native or fluent speakers of English in English in the year prior to the start of this semester.

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

2. Prior to this semester, I tried to speak English to my instructor outside of class

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

3. Prior to this semester, I tried to speak English to friends who are native or fluent speakers of English

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

4. Prior to this semester, I tried to speak English to classmates

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

5. Prior to this semester, I tried to speak English to strangers whom I thought could speak English

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

6. Prior to this semester, I tried to speak English to service personnel (e.g. bank clerk, cashier)

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

For each of the items below, choose the response that corresponds to the amount of time you estimate you spent on average doing each activity in English prior to this semester

7. Watching English language television

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

8. Reading English language newspapers

a. never
b. a few times a year
c. monthly
d. weekly
e. daily
f.

9. Reading novels in English

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

10. Listening to songs in English

a. never
b. a few times a year
c. monthly
d. weekly
e. daily
f.

11. Reading English language magazines

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

12. Watching movies or videos in English

a. never
b. a few times a year
c. monthly
d. weekly
e. daily

13. List any other activities that you commonly did using English prior to this semester.

_____
_____
_____
_____
_____
_____

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

3. Material

A. Word Cards Sample

elucidate

يشرح بالتفصيل
explain
verb
/ɪˈluːsɪdeɪt/
I shall try to *elucidate* what I believe the problems to be

gargantuan

ضخم

enormous

adjective

/gɑːˈgantjʊən/

the fat man have a gargantuan appetite

expeditious

بسرعة
speedy
adjective
/ˌekspəˈdɪʃəs/
the detective made an *expeditious* investigation of the crime

Appendix 3. Material
B. Word Card Movement Recording Sheets

First week

Name…….

Word	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Session 9	Session 10	Session 11	Session 12	Session 13	Session 14	Session 15
	Number of section														
1. Vicissitude															
2. Proprietorship		..	..		..										
3. Postulation															
4. Contentious		..	.	.	.										
5. Ingenious															
6. New-fangled		.	.		.										
7. Perceptive		.	.		.										

8. Concur															
9. Shrewd															
10. Elucidate															
11. Trespass															
12. Contrive															
13. Detestable															
14. Banish															
15. Pertinent															

a. Second week

Name…….

Word	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Session 9	Session 10	Session 11	Session 12	Session 13	Session 14	Session 15
	Number of box														

1. Vicissitude															
2. Proprietorship															
3. Postulation															
4. Contentious															
5. Ingenious															
6. New-fangled															
7. Perceptive															
8. Concur															
9. Shrewd															
10. Elucidate															
11. Trespass															
12. Contrive															
13. Detestable															
14. Banish															
15. Pertinent															
16. Adroit															

17. Stupendous														
18. Misgiving														
19. Gargantuan														
20. Irksome														
21. Mystification														
22. Ebullient														
23. Idiosyncratic														
24. Acrid														
25. Obstinate														
26. Colossal														
27. Bonny														
28. Apprehensive														
29. Brittle														
30. Unequivocally														

b. Third week

Name…….

Word	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Session 9	Session 10	Session 11	Session 12	Session 13	Session 14	Session 15
	Number of box														
1. Vicissitude															
2. Proprietorship		..	..		..										
3. Postulation															
4. Contentious		..	.	.	.										
5. Ingenious															
6. New-fangled		.	.		.										
7. Perceptive		.	.		.										
8. Concur		..		..											
9. Shrewd															

10. Elucidate															
11. Trespass															
12. Contrive															
13. Detestable															
14. Banish															
15. Pertinent															
16. Adroit															
17. Stupendous															
18. Misgiving															
19. Gargantuan															
20. Irksome															
21. Mystification															
22. Ebullient															
23. Idiosyncratic															
24. Acrid															
25. Obstinate															

26. Colossal															
27. Bonny															
28. Apprehensive															
29. Brittle															
30. Unequivocally															
31. Grievous															
32. Cogent															
33. Toil															
34. Tedious															
35. Expeditious															
36. Astound															
37. Veritably															
38. Feeble															
39. Afflicted															
40. Sluggishly															
41. Petrify															

42. Lucid															
43. Perseverance															
44. Folly															
45. Rambunctious															

Appendix 3. Material
C. Graded Readers and Comprehension Questions

Session1

Table target word percentage, number of CQs and key answer of FREE CULTURE in session 1								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*postulation (n)*	*assumption*	*1*	*b*	*postulation*
				contentious (adj)	*argumentative*	*2*	*b*	*proprietorship*
				proprietorship (n)	*ownership*	*3*	*b*	*ingenious*
				vicissitude (n)	*change*	*4*	*b*	*vicissitude*
				ingenious (adj)	*clever*	*5*	*b*	*contentious*
				banish (v)	*ban*	*6*	*a*	*banish*

FREE CULTURE

Session 1

Read the following text and carefully answer the questions.

PREFACE

At the end of his review of my first book, /Code: And Other Laws of Cyberspace/, David Pogue said that unlike actual law, it is difficult to punish who use Internet software harmfully. You may think like Pogue, who was sceptical of the postulation in the book, that software, or "code," functioned as a kind of law might be harmful. Pogue review suggested that if life on the Internet got bad, we could always flip a switch and be back home. I am doubtful about that. My postulation in Free Culture is that battles now rage regarding life on-line have fundamentally affected "people who aren't online." There is no switch that will protect us from the Internet's effect. The postulation in this book is not much about the Internet itself. It is instead about the contentious issues resulting from the Internet.

The Internet is a part of our tradition and it is the way our culture gets made. We come from a tradition of "free culture"- as in "free speech," "free markets," "free trade," "free enterprise," and "free elections". A free culture should protect those who create or invent and banishes the greedy and powerful capitalists from the market. It does this by accepting intellectual property rights so that creators can claim proprietorship of their inventions. A free culture is not a culture without proprietorship.

The opposite of a free culture is a "permission culture " a culture in which creators get to create only with the permission of creators from the past. If we understood these 1- vicissitudes of culture, I believe we would resist it. The 2- contentious story I tell here will trouble you. Given that the vicissitudes I describe, affect the value that both sides of our political culture consider fundamental. We saw how people from the right and left parties were angry as the Federal Communications Commission (FCC) considered changes in media proprietorship rules. Thousands of people are unwilling to accept the vicissitudes of rules by the FCC.

The concentration of power should be banished. The spread of power through local control is the greatest expression of democracy. This idea is an element of the contentious issue of Free Culture. My focus is not just on the concentration of power produced by concentration in proprietorship, but more importantly, my focus is on the concentration of power produced by a radical change in law.

The inspiration for the title and for many of the contentious issues comes from the work of the ingenious Richard Stallman and the Free Software Foundation. Indeed, as I reread Stallman's work, I realize that all of the theoretical insights I develop here are insights from the ingenious Stallman. Readers will think that the postulation of this work is "merely" transmitted from Stallman work. The postulation of the work of a lawyer is always transmitted from someone else, and I need to remind you of a culture about claiming full proprietorship.

Like the ingenious Stallman, I defend that tradition on the basis of values. Like the ingenious Stallman, I believe those are the values of freedom. A free culture can be our future if we do not accept the loss of freedom as a vicissitude of life. Like the ingenious Stallman's arguments for free software, a contentious issue about free culture is hard to understand. A free culture is not a culture that banishes property or banishes artists. The free culture is a balance between disorder and control. The fact that many accept such extremism as a vicissitude of fate is what I fear about our culture. The book’s goal is to banish extremism.

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1. The book general assumption is …….

a) life in the internet does not affect people in real life
b) Internet is the most fundamental part of modern life
c) life is safer without the internet

2. In Free Culture ownership of inventions ……

a) is equally distributed between generations of inventors
b) is guaranteed for original inventors but is not for follow up inventors
c) is not guaranteed to anyone

3. Stallman was described by the writer as…….

a) fast
b) very clever
c) good reader

4. Culture and law are…….. through time

a) fixed
b) changing
c) progressing

5- The issues raised by the writer in this book are ….. ..…..

a) common and agreed upon by majority of people
b) are argumentative and people may have different opinions about it
c) easy to understand

6- According to the writer a free culture should supports and protects those who create or invent and …… the greedy and powerful capitalists from the market.

a) dismiss
b) give a chance to
c) sue

Session 2

Table target word percentage, number of CQs and key answer of FREE CULTURE in session 2								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*perceptive (adj)*	*knowledgeable*	*1*	*c*	*perceptive*
				elucidate (v)	*demonstrate*	*2*	*b*	*trespass*
				new-fangled (adj)	*modern*	*3*	*a*	*elucidate*
				concur (v)	*agree*	*4*	*c*	*concur*
				trespass (v)	*infringe*	*5*	*a*	*new-fangled*
				shrewd (adj)	*intelligent*	*6*	*b*	*shrewd*

FREE CULTURE

Session 2

Read the following text and carefully answer the questions.

INTRODUCTION

In 1903, the Wright brothers elucidate that a heavier-than-air vehicle could fly. The Wright brothers' perceptive claims created an explosion of interest in this new-fangled technology. A ton of perceptive innovators began to build upon it and concur in the determination to invent. At the time the of the new-fangled airplanes, American law held that a property owner owned not just the surface of his land, but all the land below and all the space above. For many years, perceptive scholars had puzzled about how best to interpret this law. Does this mean you could prosecute airplanes when they fly on the sky over your land? For the first time, this principle of American law mattered to perceptive legal thinkers.

When the Causbys started losing chickens because of low-flying military aircraft (the scared chickens flew into the animal shelter walls and died), the Causbys filed a lawsuit saying that the government was trespassing on their land. The airplanes, of course, never intrude into the surface of the Causbys' land. But the land was trespassed on its sky then the government was trespassing on their property. The Supreme Court concurred to hear the Causbys' case.

However, the Court concurred that the old law of property ownership need to be changed. Every flight would be subject to countless trespass suits if the law continues to be the same. Common sense revolts at ideas that were considered as shrewd in one age then come to nothing in another. Or at least, this is how history has elucidated things happen.

Other farmers would find it hard to unite and stop the new-fangled airplanes from trespassing their property. The shrewd Wright brothers' invention spread and farmers found themselves surrounded by their technology. They could stand on their farms and protest but their "private interest" would not be allowed to defeat an obvious public gain.

The shrewd Edwin Howard Armstrong work in the area of radio technology was the most important. Perceptive Armstrong invented important technologies that advanced our understanding of radio. Four copyrights in inventions were issued to Armstrong for his most new-fangled invention FM radio. However, consumer radio had been AM radio even if FM radio is clearer.

In 1935, Armstrong elucidated the new-fangled technology. He tuned his radio dial across a range of AM stations. The radio fell totally silent, and then with a clarity no one else in that room had ever heard the sound of an announcer's voice. But at the time of his invention, the shrewd Armstrong was working for RCA. RCA was the dominant player in the dominant AM radio market.

RCA's president, David Sarnoff was eager about the shrewd Armstrong discovery. But when Armstrong elucidated his invention, Sarnoff discovers that this invention threatened RCA's AM empire. Therefore, RCA did not support FM radio and launched a campaign to destroy it. RCA began to use its power and with the government to stop FM radio's distribution generally.

AT&T strongly concurred with this war, because the loss of FM stations would mean radio stations would have to buy wired links from AT&T. The spread of FM radio was thus blocked by the RCA. After a long battle, RCA offered a settlement so low for Armstrong. Defeated and broken, in 1954 Armstrong committed suicide. This is how the law sometimes shockingly works.

Most people would concur that powerful interest too often makes use of its influence within the government to be protected. As history elucidated ideas of this subtle corruption of our political

process. RCA had what the Causbys did not: the power to prevent the effect of technological change.

Now that you read the test please answer the following.

Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1. Wright brothers and Armstrong are described in the text as ……

    a) wise people
    b) patient scholars
    c) knowledgeable inventors

2. Causbys wanted to gain their right of banding plains from ……
    a) landing in their farms
    b) entering the sky of the farm without permission
    c) flying around the farm

3. History …….. how point of views and opinion are subject to change.
    a) explains to us in detail
    b) reject
    c) predict

4. AT&T …….. with RCA banding FM radio because the loss of FM stations would mean radio stations would have to buy wired links from AT&T.

    a) discussed
    b) disagree
    c) agree

5. The …….. invention of airplanes brings with it new rules and regulations.

    a) brand new
    b) unpredicted
    c) surprising

6. Armstrong was very…… that he was able to discover FM radio.

    a) supported
    b) smart
    c) challenged

Session 3

Table target word percentage, number of CQs and key answer of FREE CULTURE in session 3								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*vicissitude (n)*	*change*	*1*	*a*	*vicissitude*
				new-fangled (adj)	*modern*	*2*	*b*	*new-fangled*
				ingenious (adj)	*clever*	*3*	*a*	*proprietorship*
				contrive (v)	*construct*	*4*	*c*	*concur*
				proprietorship (n)	*ownership*	*5*	*a*	*contrive*
				concur (v)	*agree*	*6*	*a*	*ingenious*

FREE CULTURE

Session 3

Read the following text and carefully answer the questions.

In a very short time, everyone agree that the Internet has become part of everyday life and it has created noticeable vicissitudes. Some of these vicissitudes are technical in nature, like faster communication, gathering data with lowered cost and so on. These new-fangled technical vicissitudes are not the focus of this passage. Everyone concur that they are important but they are not well understood. This passage is about how culture is contrived by the Internet.

My claim is that the Internet has contrive unrecognized vicissitudes in our culture. That new-fangled change will transform old tradition. Most, if they recognized these vicissitudes and their values, would reject them. We can detect a sense of this change by distinguishing between commercial and non-commercial culture.

For just about the whole of our tradition, non-commercial culture was essentially unregulated. Of course, if your creation had bad values then most of us will concur that the law might intervene. But the law was never directly concerned with spread of this form of culture, and it left this culture "free."

The law protected the rights of creators by granting them exclusive rights to their ingenious creation. This is of course, important part of creativity and culture. However, for the first time in our tradition, the ordinary ways in which individuals contrive culture fall within the reach of the regulation of the law, which has expanded to draw within its control ingenious inventions.

The new-fangled technology did not preserve the balance between uses of our culture and the consequence is that we are less a free culture, more a permission culture. This change gets justified as necessary to protect commercial work. But this protectionism is to protect certain forms of business. It is the story of RCA and the ingenious Armstrong; it is the dream of the Causbys.

The Internet has the power to contrive new culture. Digital technologies could produce an active market that could include a much wider and more diverse range of creators with ingenious inventions. Those creators could earn more on average from this new-fangled system than creators do today.

These modern-day equivalents of the early twentieth-century radio or nineteenth-century railroads are succeeding in their plan to remake the Internet before the Internet remakes them. They are using their power to get the law to protect them against more vibrant new-fangled technology for building culture. They

The battles over proprietorship in the Internet seem remote to most. To the few who follow them, they seem mainly about "piracy," property protection and war against the technologies of the Internet. It is Motion Picture Association of America (MPAA) president battle about the rule of law and respect for proprietorship. To know which side to take in this war we need only decide whether we are for proprietorship or against it.

I am a believer in proprietorship, and I believe that the law, properly should punish "piracy,". However, my fear is that the war to rid the world of Internet "pirates" will also rid our culture of values and traditions.

Professor Netanel powerfully argues, that Internet law of proprietorship if properly balanced will protect creators against private control. Our tradition support creators to extend our culture. Yet, the law's response increased the effective regulation of ingenious creation in the internet.

We have contrived a kind of cultural nobility. But it is nobility of any form that is alien to our tradition. There will be great harm to our tradition if the struggle allowed to continue

unchecked. We must concur that we need to understand the source of this war. We must concur to resolve it soon.

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1. The use of Internet in everyday life create …….. in our culture.
   a. changes
   b. challenges
   c. inventions

2. The ……. Technology brings with it different challenges to protect traditions and ownership.

   a. evil
   b. modern
   c. difficult

3. Few people are concerned with ……. Of the Internet because there is not enough awareness of its effect in our culture.

   a. ownership
   b. properties
   c. technology

4. The writer insists on readers to ……. With him about the importance of fighting the battel against those who wants to protect their ownership against new invention.

   a. argue
   b. judge
   c. agree

5. Technology and Internet should serve to …….. our culture not to destroy it.

   a. build
   b. limit
   c. save

6. The writer suggest that law does not really grant the right of …….. creation. In fact, the law help sell those exclusive rights in a commercial marketplace.

   a. clever
   b. fake
   c. false

Table target word percentage, number of CQs and key answer of FREE CULTURE in session 4								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*perceptive (adj)*	*knowledgeable*	*1*	*a*	*pertinent*
				shrewd(adj)	*intelligent*	*2*	*c*	*perceptive*
				pertinent (adj)	*relevant*	*3*	*a*	*contentious*
				detestable (adj)	*awful*	*4*	*b*	*detestable*
				contentious (adj)	*argumentative*	*5*	*a*	*trespass*
				trespass (v)	*infringe*	*6*	*b*	*shrewd*

FREE CULTURE

Session 4

Read the following text and carefully answer the questions.

The contentious issue I am going to address is not much about the Internet itself, it is instead pertinent to the war of the Internet and tradition. There are contentious changes of our culture caused by the Internet. Tradition is the way our culture gets made and we come from a tradition of "free culture" not "free" as in "free drink" but contentious "free" as in "free speech," "free markets," and "free trade".

This detestable war is about property and trespassing on this property. The property of this detestable war is not as tangible as the Causbys'. The ideas pertinent to owning property and trespassing on this property are as obvious to most as the Causbys' claim about the importance of their farm was to them. Most of us take for granted the powerful claims that the owners of "intellectual property" now assert. Like the Causbys, we treat these claims as obvious. And hence, the Causbys, object when a new technology trespasses upon this property. It is as obvious to us as it was to them that the new technologies of the Internet are "trespassing" upon legitimate claims of "property." It is as obvious to us as it was to them that the law should intervene to stop this trespassing. And thus, when enthusiasts and perceptive technologists defend their shrewd Armstrong or shrewd Wright technology, most of us are simply unsympathetic.

Common sense does not revolt. Unlike in the case of the unlucky Causbys, common sense is on the side of the property owners in this war. Unlike the shrewd Wright brothers, the Internet has not inspired a revolution on its side. My hope is to push this common sense along. I have become increasingly amazed by the power of this perceptive intellectual property and, more pertinent, its power to disable critical thought by policy makers and citizens.

Is the radical and shift away from our tradition of free culture an instance of America correcting a detestable mistake from its past, as we did after a bloody and a detestable war with slavery, and as we are slowly doing with inequality? Or is the radical shift away from our tradition of free culture yet another example of a political system captured by a few powerful interests? Does common sense actually believes in these extremes? Or does common sense stand silent in the face of these extremes as with shrewd Armstrong versus RCA, the more powerful side ensured the more powerful view? I want to express what perceptive people would think. My own views pertinent to these questions are resolved.

The struggle that rages just now centres on two ideas: "piracy" and "property." My method is not the usual method of a perceptive academic and shrewd inventors. I don't want to throw you into a complex argument, however natural that is for the weird sort we perceptive academics have become. Instead I begin in each part with a collection of stories that set a context within which these apparently simple concentration can be more fully understood.

Rather than understanding the changes the Internet might permit, and rather than taking time to let "common sense" resolve how best to respond to pertinent issues, we are allowing those most threatened by the changes to use their power to change the law and more importantly, to use their power to change something fundamental about who we have always been. We allow it because the interests most threatened are among the most powerful players in our detestable process of making law. This book is about the contentious story of one more consequence of this form of corruption.

Now that you read the test please answer the following.

Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1. The writer is concerned about issues …….. the war of the use of the Internet and tradition.

   a. related to
   b. against
   c. caused

2. The writer wants the reader to join the …… technologist who defend the Wright brothers and Armstrong.

   a. afraid
   b. distinguished
   c. knowledgeable

3. In a number of places in this text the writer describes the issues caused by the Internet and technology as ……..

   a. controversial
   b. evil
   c. obvious

4. The …….. war the writer want readers to fight is not only about property, but also about creativity, tradition and culture.

a. ugly
b. awful
c. unavoidable

5. The Causby's problem was trying to find a law that defend them from airplane …….. the sky above their land.

a. infringing
b. bombing
c. polluting

6. The Wright brothers and Armstrong are often described as……. In the text.

a. helpful
b. smart
c. lucky

Session 5

Table target word percentage, number of CQs and key answer of FREE CULTURE in session 5								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*pertinent (adj)*	*relevant*	*1*	*a*	*postulation*
				detestable (adj)	*awful*	*2*	*c*	*detestable*
				banish (v)	*ban*	*3*	*b*	*pertinent*
				elucidate (v)	*demonstrate*	*4*	*a*	*banish*
				postulation (n)	*assumption*	*5*	*a*	*elucidate*
				contrive (v)	*construct*	*6*	*b*	*contrive*

FREE CULTURE

Session 5

Read the following text and carefully answer the questions.

CONCLUSION

This book's postulation is about the detestable effect of the Internet beyond the Internet itself. The Internet has created an important change that is pertinent to the creation of our culture. That change will radically transform a tradition that is very old and contrive new and different tradition. If people recognized this change pertinent to their values they would reject it. Yet most people do not even understand the postulation about the change that the Internet has introduced.

We can look a sense of change caused by the Internet by elucidating the law's regulation of commercial and non-commercial culture. By "commercial culture" I mean that part of our culture that is produced and sold. By "non-commercial culture" I mean all the rest. For example, when old men sat around parks or on street corners telling stories that kids and others consumed, that was non-commercial culture. When Noah Webster published his "Reader," or Joel Barlow his poetry, that is pertinent to commercial culture.

At the beginning of our history non-commercial culture was banished. Stories that had detestable effect pertinent to immoral issues, or songs that disturbed the peace they were banished by the law. But the law was never directly concerned with the postulation about the bad effect of the Internet nor in the creation or spread of this form of culture. The law maker did not banish the non-commercial culture nor did they contrive law for it.

For the first time in our tradition, the ordinary ways in which individuals contrive and share culture fall within the reach of the regulation of the law. These regulations of law have expanded to contrive within its control a vast amount of culture and work. Protectionism that justifies the changes, elucidated previously are protectionism to protect certain forms of business.

For the Internet has released an amazing possibility for many to participate in the process of building a culture that reaches far beyond local boundaries. That power has changed the marketplace for contriving culture generally, and that change in turn threatens established content industries.

The postulation I am attempting to address is that the Internet is thus the industries that made the distribution of content in the twentieth century. As I elucidated previously the battles over ownership of Internet property and the Internet seem not relevant to many. To elucidate, the few who follow ownership rules, they seem mainly about a much simpler set of questions--whether "piracy" will be banished, and whether "property" will be protected. It is simply a postulation about the "war" that has been waged against the technologies of the Internet.

I believe that "piracy" is wrong, and that the law, properly tuned, should banish who is involved in "piracy". But those simple beliefs mask a much more fundamental question and a much more dramatic change.

Academia elucidates that culture around us is, of course, often granted--but it is not often granted to the critical or the independent. Is it not about the "centrality of technology" to ordinary life. It is not a morality tale. Nor is it a call to war against an industry. It is instead an effort to understand a hopelessly detestable war inspired by the technologies of the Internet but reaching far beyond its code. And by understanding this detestable battle, it is an effort to map peace.

There is no good reason for the issues pertinent to this struggle around Internet technologies to continue. There will be harm to our tradition if it can continue unchecked. We must come to understand the source of this detestable war. We must resolve it soon.

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1. The book is proposing a …… about the effect of the Internet on our culture and traditions.

a) theory
b) fact
c) message

2. The writer wants the readers to understand that the ……. battle he is explaining about is an effort to map peace.

a) loosing
b) imaginary
c) awful

3. The Internet has created an important change that …….. the creation of our culture.

a) did not affect
b) is related to
c) prevents

4. The writer suggests that the law should do more than …….. piracy and protecting property.

a) getting rid of
b) punishing
c) preventing

5. The writer …….. how powerful businessmen will use their power to get the law to protect them against new inventions.

a) explains in detail
b) summarizes
c) fails to explain

6. For the Internet has released an amazing possibility for many to participate in the process of ……. a culture that reaches far beyond local boundaries.

a) discovering
b) building
c) investigating

Session 6

Table target word percentage, number of CQs and key answer of A SCANDAL IN BOHEMIA in session 6								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
7	35	*700*	*7*	*adroit (adj)*	*skilled*	*1*	*b*	*stupendous*
				stupendous (adj)	*amazing*	*2*	*a*	*adroit*
				misgiving (n)	*doubt*	*3*	*b*	*irksome*
				gargantuan (adj)	*huge*	*4*	*a*	*ebullient*
				irksome (adj)	*annoying*	*5*	*c*	*misgiving*
				mystification (n)	*puzzlement*	*6*	*c*	*mystification*
				ebullient (adj)	*enthusiastic*	*7*	*b*	*gargantuan*

A SCANDAL IN BOHEMIA

Session 6

Read the following text and carefully answer the questions.

Sherlock Holmes is the most stupendous reasoning and adroit observing machine that the world has seen, but he is not a lover. He never spoke emotions as they are irksome to him. Emotions are irksome things for the adroit observer, excellent for drawing the veil from men's motives and actions. Love and passion might express misgivings about all his mental results. Grit in a sensitive instrument, or a crack in one of his own gargantuan power lenses, would not be more disturbing and irksome than emotion in a nature such as his. And yet there was but one woman to him, and that woman was the stupendous Irene Adler. To Sherlock Holmes she is always THE woman. In his eyes, she eclipses the whole of her sex.

My own complete happiness and ebullient feelings to the home centred interests kept me away from the man who first finds himself master of his own establishment in solving cases with mystifications. Holmes who finds marriage and family irksome, and gargantuan commitment remained in our lodgings in Baker Street. He was still, as ever, deeply ebullient by the study of mystifications, and occupied his gargantuan faculties and stupendous powers of adroit observation in following out those clues, and clearing up those mystifications.

One night, I was returning from a journey to a patient, when my way led me through Baker Street. I was captured with an ebullient desire to see Holmes again, and to know how he was employing his adroit observations. He was ebullient upon the scent of some new mystifications. I rang the bell and was shown up to the chamber.

He wasn't ebullient to see me. He seldom was; but he was glad, I think! He stood before the fire and said. "Marriage suits you," he remarked. "I think, Watson, that you have put on seven and a half pounds since I saw you."

"Seven!" I answered.

"And in practice again, I observe. You did not tell me."

"My dear Holmes," said I, "this is stupendous. It is true! I fail to see how you work it out and how you seem to have no misgivings about it"

He chuckled to himself and rubbed his long, nervous hands together. He said:

"If a gentleman walks into my rooms smelling of medicine, with a black mark of nitrate of silver upon his finger, I must be irksome, indeed, if I do not pronounce him to be an active member of the medical profession. You see, but you do not observe. The gargantuan distinction is clear. Since you are interested in my adroit observation, you may be interested in something more stupendous".

He threw over a sheet of thick, pink – coloured note

The note was undated, and without either signature or address.

"A gentleman who desires to consult you upon mystification as your recent services to one of the royal houses of Europe have shown with no misgiving, that you can be trusted. Be in your chamber at eight o'clock and do not think anything is wrong if your visitor wears a mask." It said

"What do you imagine that it means?" I said

"I have no data yet! What do you deduce from it?"

I carefully examined the writing, and the paper upon which it was written.

"Such paper could not be bought under half a crown a packet. It is peculiarly strong and stiff."

"Peculiar – that is the very word," said Holmes. "this clears the misgiving that is not an English paper at all. Hold it up to the light."

I did so, and saw a large "E" with a small "g," a large "G" with a small "t" .

"What do you make of that?" asked Holmes.

"The name of the maker, there is no misgivings about that."

"Not at all. The 'G' with the small 't' stands for 'Gesellschaft,' which is the German for 'Company.' Eglow, Eglonitz – here we are, Egria. It is in a German – in Bohemia, not far from Carlsbad. 'Remarkable as being the scene of the death of Wallenstein.
Ha, ha, my boy, what do you make of that?" His eyes sparkled, and he sent up a gargantuan blue triumphant cloud from his cigarette.

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1- For Sherlock Holmes Irene Adler is…….
   1. a friend
   2. an amazing woman
   3. his sister

2- Sherlock Holmes is the most …….. observing machine that the world has seen
   1. skilled
   2. famous
   3. important

3- Holmes finds marriage and family …….
   1. important
   2. annoying
   3. just unnecessary commitment

4- Dr Watson, the narrator, is ……. To see Holmes again
   1. enthusiastic
   2. worried
   3. not interested

5- Dr Watson fail to see how Holmes work things out and how he seems to have no……about it.
   a idea
   b clue
   c doubt

6- The note indicates that a gentleman desires to consult Holmes upon ……
   1. sickness
   2. a murder
   3. puzzlement

7- After knowing where the note came from, Holmes was happy and he sent up …….. blue triumphant cloud from his cigarette.
   1. a huge
   2. an amazing
   3. a small

Session 7

Table target word percentage, number of CQs and key answer of A SCANDAL IN BOHEMIA in session 7								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*idiosyncratic (adj)*	*distinctive*	*1*	*c*	*idiosyncratic*
				acrid (adj)	*sharp to taste or smell*	*2*	*a*	*acrid*
				misgiving (n)	*doubt*	*3*	*c*	*obstinate*
				obstinate (adj)	*stubborn*	*4*	*b*	*misgiving*
				colossal (adj)	*very large*	*5*	*a*	*mystification*
				mystification (n)	*puzzlement*	*6*	*b*	*colossal*

A SCANDAL IN BOHEMIA

Session 7

Read the following text and carefully answer the questions.

" The paper was made in Bohemia," I said.
"Do you note the idiosyncratic construction of the sentence'. This is not a mystification to me, it is the German who is so impolite to his verbs"
And here he comes, if I am not mistaken, to resolve all of our misgivings."
Glancing out of the window Holmes said: " there was acrid smoke coming from an expensive carriage. There's money in this case!"
"I think that I had better go, Holmes."
"Not a bit, Doctor. Stay where you are."
"But your client – "
" Don't be very obstinate."
In a while there was a loud tap and an acrid smell coming from the door.
"Come in!" said Holmes.
A man entered with idiosyncratic height and with acrid smell of odor of burning cigarettes. His dress was rich with a richness. He wears idiosyncratic heavy bands of wool fabric were slashed across the sleeves and fronts of his colossal double breasted coat. Boots which were trimmed at the tops with brown fur, completed the idiosyncratic appearance. He carried a colossal hat in his hand, while he wore a black mask.
"You had my note?" with a strongly marked German accent with and an acrid breath coming from his mouth.
"I should communicate with you alone." Said the strange man
I rose to go, but obstinate Holmes caught me by the wrist and pushed me back into my chair. "It is both, or none," said my obstinate friend.
The man shrugged his broad shoulders. "my mystification is of such weight it may have an influence upon European history."
"I promise," said Holmes. "And I."
"You will excuse this mask," continued our visitor. "I have to be mysterious to you"
"I was aware of it," said Holmes.
"my mystification might cause a colossal scandal and seriously compromise the kings of Bohemia."
"If your Majesty would agree to state your case," he remarked, "I should be better able to advise you." Said Holmes
The colossal man sprang from his chair and became angry. Then, with a gesture of desperation, he throws his mask to the ground. "You are right," he cried and the acrid smell of odor of burning cigarettes comes out from him; "it seems with no misgiving that you know that I am the King. Why should I attempt to keep it as a mystification?"
" I was aware that I was addressing Wilhelm Gottsreich Sigismond von Ormstein, Grand Duke of Cassel – Felstein, and King of Bohemia but you were obstinate to be mysterious"
our visitor passes his hand over his high white idiosyncratic forehead, "you can understand that I am not accustomed to doing such business in my own person, but it is a delicate matter."
Five years ago, I made the acquaintance of the famous Irene Adler. The name is with no misgivings familiar to you."
As I understand, you have a mystification with this person?
"It is marriage and no legal papers or certificates?"
"Then I fail to follow your Majesty. How can she prove any documents' authenticity?"
"We were both in a photograph! She is obstinate and won't let me have it"

“ It must be bought.”
“She will not sell.”
“With no misgivings, stolen!”
“Her house was searched twice There has been no result.”
“ And what does she propose to do with the photograph?”
“To ruin me! I am about to be married to the second daughter of the King of Scandinavia. You may know the principles of her family. A shadow of a misgiving as to my conduct would cause a colossal scandal”.

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1. Holmes visitor’s clothes looks ……

a) normal
b) beautiful
c) distinctive

2. The German visitors has a …….. smell of cigarette

a) sharp
b) nice
c) strange

3. Holmes was …… he did not let Dr Watson go.

a) mad
b) scared
c) stubborn

4. The German man has no ……. That Holmes knows Irene Adler.

a) idea
b) doubt
c) expectations

5. The king threw the mask when he realizes that he can’t keep his identity as a ……

a) secret
b) nobleman
c) stranger

6. The German man was …… in size

a) tiny
b) large
c) average

Session 8

Table target word percentage, number of CQs and key answer of A SCANDAL IN BOHEMIA in session 8								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*bonny (adj)*	*beautiful*	*1*	*a*	*ebullient*
				apprehensive (adj)	*anxious*	*2*	*c*	*obstinate*
				ebullient (adj)	*enthusiastic*	*3*	*c*	*colossal*
				obstinate (adj)	*stubborn*	*4*	*b*	*brittle*
				colossal (adj)	*very large*	*5*	*b*	*apprehensive*
				brittle (adj)	*breakable*	*6*	*c*	*bonny*

A SCANDAL IN BOHEMIA

Session 8

Read the following text and carefully answer the questions.

“And Adler?” said Holmes
“Threatens to send them the photograph. And she will! She is obstinate. She has an obstinate mind and a soul of steel. She has the face of the bonniest of women, and the mind of the most obstinate of men. She has said that she would send it on the day of engagement next Monday.” Said the king
“Oh, then we have time,” said Holmes with an ebullient manner. “We should not feel apprehensive. Holmes took the address from the king.
“Was the photograph a cabinet?” asked Holmes
“It was. I will be all apprehensive. I might look like I have a hard as an iron soul but I am as brittle as glass, this can cause a colossal scandal “
“Please don’t feel apprehensive and I trust that we shall soon have some good news”
Holmes turned to me and said, “I will see you tomorrow at three o’clock”

II.
At three o’clock precisely I was at Baker Street, but Holmes was late. I was shuffling brittle pages of his old books while I was waiting. The nature of the case is as brittle as glass, I am ebullient to help. However, so accustomed was I to Holmes colossal success and obstinate personality I had doubts that he might fail.
It was four o’clock before the door opened, and a man with dirty clothes, brittle hair, walked into the room. I knew my friend’s powers in disguises, I was certain that it was Holmes. He vanished into the bedroom and emerged looking clean and respectable. An ebullient laugh comes from Holmes and he said
“It’s quite funny. I am sure you will be ebullient if you know how I spent my morning”
“ I suppose that you have been watching the house, of bonny Adler.”
“Quite so! “ He said that with ebullient manner.
“I left the house this morning in the character of a groom out of work. Be a horsey man and you will know all that there is to know. I soon found the villa, with a colossal garden at the back. Near to the villa’s entrance a colossal sitting room on the right side, well furnished, with colossal windows almost to the floor. I examined it but without noticing anything significant. I then went to the stables and lent the men a hand in rubbing down their horses to find information as about Adler”
“And what of her?” I asked.
“She is the most delicate thing under a hat on this planet, she looks like if she has brittle bones Yet she is obstinate. Said one of the stable men. She has only one male visitor. He is bonny and comes once or twice a day. He is Mr. Norton, her lawyer!”
“She had probably transferred the photograph to Mr. Norton keeping?”
“I was still thinking about it when a gentleman sprang out of a cab and was in a hurry. He was a remarkably bonny man, evidently the man of whom I had heard. He was in the house and I could catch glimpses of him in the windows talking, and waving his arms. It seems he was apprehensive. Presently he emerged, looking even more apprehensive than before. As he stepped up to the cab, he shouted, ‘Drive very fast! ‘
“I wanted to follow him when up the lane came a carriage. It hadn’t pulled up before Adler shot out the doors. I only caught a glimpse of her at the moment, but she was a bonny woman, with a face that a man might die for. “

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1 Holmes is …… about solving cases

a enthusiastic
b careless
c selfish

2 Adler could do what she threats the king for because she is ……

a) stupid
b) unkind
c) stubborn

3 Adler's sitting room has …….. windows almost to the floor.

a) very few
b) very ugly
c) very large

4 While Dr Watson was waiting to Holmes he was shuffling……. Pages of his old books.

a) old
b) thin and almost breakable
c) heavy

5 Holmes could catch glimpses of the gentleman in the windows talking, and waving his arms and he seems to be ……..

a) happy
b) extremely worried
c) funny

6 According to Holmes, Adler was a ……woman, with a face that a man might die for

a) short
b) powerful
c) beautiful

Session 9

Table target word percentage, number of CQs and key answer of A SCANDAL IN BOHEMIA in session 9								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*idiosyncratic (adj)*	*distinctive*	*1*	*b*	*unequivocally*
				acrid (adj)	*sharp to taste or smell*	*2*	*a*	*irksome*
				gargantuan (adj)	*huge*	*3*	*a*	*gargantuan*
				irksome (adj)	*annoying*	*4*	*c*	*idiosyncratic*
				adroit (adj)	*skilled*	*5*	*a*	*adroit*
				unequivocally (adv)	*certainly*	*6*	*a*	*acrid*

A SCANDAL IN BOHEMIA

Session 9

Read the following text and carefully answer the questions.

Adler went in the cab 'The Church of St. Monica' she cried,
"I unequivocally followed her. I took a cab and irksome driver asked twice the tickets money, but I jumped in before we agree in the cost. "
'The Church of St. Monica,' said I
"My irksome cab driver drove fast. The cab wheels made an acrid smoke when I arrived. I paid the irksome man, who was not happy with the payment, and hurried into the church. There were three standing in front of the altar. Suddenly the three people face round to me, and the idiosyncratic Mr Norton came running.
"Thank God,' he cried 'You'll do. Come! Come! only three minutes, or it won't be legal. "
"I was half – dragged it was irksome! I found myself assisting the marriage of Miss Adler, to Mr Norton. It was unequivocally the most idiosyncratic position. It seems that they need a witness, and that my lucky appearance saved them from the irksome search for a best man."
"This is a very idiosyncratic turn of affairs," said I; "and what then?"
"Well, it looked as if the pair drove away in different directions and so I needed adroit measures on my part and to make my own arrangements."
"Which are?"
“a bath to get rid of my acrid bad smell from the stable then some food”.
"Doctor, I need an adroit helper, can you cooperate."
"I am unequivocally your man"
"In two hours, we must be on the scene of action. Miss Adler, or Madame, rather, unequivocally returns from her drive at seven. We must be at her house"
"And what then?"
"There will probably be some small idiosyncratic situations. Do not join in it. You are to stay close to an open window. And when I raise my hand, you will throw into the room what I give you to throw and raise the cry of fire."
"He took a long cigar from his pocket. "I want you to be adroit, nothing idiosyncratic, it is an ordinary plumber's smoke not a rocket with acrid smoke. When you raise your cry of fire, it will unequivocally be taken up by people. You may then walk to the end of the street, and I will re-join you"
"I have to be adroit on my new role.”
Holmes disappeared into his bedroom and returned in the character of a clergyman. He is adroit in disguising and he planned to go to Adler’s house as a clergyman.
It was a quarter past six we arrived at the gargantuan house. It was just such as I had pictured it from Holmes' adroit description, but the area appeared to be less private than I expected. There was a group of poorly dressed men with an acrid smell of smoking and well - dressed young men who were holding cigars.
Where are we to find the photograph? Asked Holmes
"It is too gargantuan for easy concealment about a woman's dress. We may take it, then, that she does not carry such a gargantuan thing about with her."
"Where, then? "
" I am inclined to think she gave it to someone. Why should she hand a gargantuan picture over to anyone else especially if can cause a scandal? It must be in her own house."
"But it has twice been burgled. How can they not have stolen a gargantuan picture?"

"I will get her to show me."
"But she will refuse."
"She will not be able to. But I see an acrid smoke from the wheels of her carriage. Now carry out my orders to the letter!

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1- Holmes found that he will …….. going to follow Adler to the church

a) quickly
b) certainly
c) slowly

2- The driver was ……. because he asked more money from Holmes

a) annoying
b) fast
c) helpful

3- The photograph Holms and Dr Watson are looking for is …….. in size.

a) large
b) small
c) medium

4- When Holmes assessed Miss Alder and Mr Norton marriage, he found himself in the most ……position

a usual
b normal
c distinctive

5- Holmes is ……. in disguising as a clergyman

a) skilled
b) speedy
c) fine

6- Holmes was planning to take a bath as he had …… smell from being in the stable

a) bad and sharp
b) mild
c) strange

Session 10

Table target word percentage, number of CQs and key answer of A SCANDAL IN BOHEMIA in session 10								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
5	25	*500*	*5*	*bonny (adj)*	*beautiful*	*1*	*c*	*bonny*
				apprehensive (adj)	*anxious*	*2*	*b*	*stupendous*
				unequivocally (adv)	*certainly*	*3*	*b*	*brittle*
				stupendous (adj)	*amazing*	*4*	*b*	*apprehensive*
				brittle (adj)	*breakable*	*5*	*a*	*unequivocally*

A SCANDAL IN BOHEMIA

Session 10

Read the following text and carefully answer the questions.

Adler's carriage pulled up to the door of the villa, a gang of men fight to open her door. In an instant bonny Adler was the centre of men fighting. Holmes dashed into the fight to protect bonny Adler as she looks as brittle as a glass in this crowd; but just as he reached her he dropped to the ground, with blood running down his face. Adler, was apprehensive and had to hurry up the steps; but she stood at the top with her stupendous figure outlined against the lights of the hall.
"Is he hurt?" asked Adler.
"No, he's alive!" cried apprehensive voices.
"Bring him in, please!" said Adler
While I still observing so that I could unequivocally see Holmes as he lay inside the villa.
I felt ashamed when I saw the bonny creature against whom I was plotting, or the stupendous grace with which she treated my injured friend. I hardened my brittle heart when I saw Holmes raise his hand and at the signal I tossed my rocket into the room with a cry of "Fire!" then the whole apprehensive crowd scream of "Fire!". Stupendous clouds of smoke curled through the room. A moment later Holmes was assuring everyone that it was a false alarm. Slipping through the apprehensive crowd I made my way to the street, and in ten minutes was happy to find my friend's arm in mine.
"You have the photograph?" I said
"I know where it is."
"And how did you find out?"
"She showed me! "
"I don't understand".
"When the fight happened, I had a little red paint in my hand. I clapped my hand to my face and looked pathetic. It is unequivocally an old trick. Then everyone become apprehensive about me".
"Stupendous skills! "
"Adler had nothing more precious to her than the brittle picture. She rushed to save her brittle photograph when she thought there is fire. I saw where she hides it. When I cried that it was a false alarm, she replaced it. I wanted to secure the photograph but it seemed safer to wait. I shall call the King tomorrow to regain it with his hands."
The next day Holmes explains to the king what happened and we went to the villa.
In the way, the king said " I wish she had been my wife! What a bonny queen she would be!"
When we arrived, a servant opened the door and said, "My mistress left England this morning"
"What?!
We rushed into the villa. The furniture was unequivocally scattered, and we found a letter. We opened it and read it together.
"DEAR MR. SHERLOCK HOLMES, You, unequivocally took me in. I had been warned against your stupendous skills months ago. I had been told that if the King employed an agent it would unequivocally be you. Yet, you made me reveal what you wanted to know. I will keep the photograph to preserve a weapon which will always secure me.
Yours, ADLER

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1- Dr. Watson, the narrator, often describes Miss Adler as…….

a) Selfish
b) Mysterious
c) Beautiful

2- Dr. Watson, the narrator, is impressed by Holmes ……. Skills

a) unusual
b) amazing
c) different

3- Holmes dashed into the fight to protect Adler as she looks ……. In this crowd

a) transparent like a glass
b) breakable like a glass
c) solid like a glass

4- Adler was …… about Holmes when she saw the blood in his face

a) careless
b) worried
c) not thinking

5- Adler said in her letter that she had been told that if the King employed an agent it would …… be Holme

a) certainly
b) not
c) with doubts

Session 11

Table target word percentage, number of CQs and key answer of METAMORPHOSIS in session 11								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*folly (n)*	*nonsense*	*1*	*b*	*folly*
				feeble (adj)	*weak*	*2*	*a*	*feeble*
				toil (v)	*work hard*	*3*	*b*	*perseverance*
				perseverance (n)	*endurance*	*4*	*a*	*rambunctious*
				tedious (adj)	*dull*	*5*	*a*	*tedious*
				rambunctious (adj)	*loud*	*6*	*c*	*toil*

Metamorphosis

Session 11

Read the following text and carefully answer the questions.

One morning, when Gregor woke from nightmares full of folly, he found himself transformed into a horrible bug. He lay on his weird back, and if he lifted his head a little he could see his brown belly. His many legs are feeble and thin compared with the size of the rest of him, waved about helplessly as he looked. "This is folly! What happened to me?" he thought. Gregor then turned to look out the window at the tedious weather. It was raining. "How about if I sleep a little bit and forget all this folly", he thought, but that was something he was toiling to do because he was used to sleeping on his right, and in his present state he is too feeble to get into that position. However hard he threw himself onto his right, he always rolled back to where he was. Trying to escape this tedious situation, he shut his eyes so that he wouldn't have to toil or look at his feeble legs.

He thought, "what a tedious career it is that I've chosen! Travelling day in and day out. Doing business like this takes more perseverance. You toil to have social life. He looked at his body, he found it covered with white spots which he didn't know what to make of; and when he toiled to feel the place with one of his feeble legs he couldn't.

He looked over at the clock "oh no!" he thought. It was half past six. The next train went at seven; he needs to catch it. Even if he did catch the train he would not avoid his rambunctious boss being angary. In fifteen years of service Gregor had never once yet been ill, feeble or late. His rambunctious boss would certainly come around with the doctor from the medical insurance company.

Soon his father came knocking, gently, but with his fist. "Gregor, Gregor", he called, "what's wrong?" And after a short while he became rambunctious: "Gregor! Gregor!" then his sister came: "Gregor? Do you need anything?" Gregor answered with perseverance as talking is hard for him. But no one recognize what's he's saying.

Gregor wanted to get out of bed but he had all those little legs continuously moving in different directions and he was unable to control them. "This is folly! " He tried to move himself forwards with perseverance, but he hit his head very hard.

He realised that he better stays in bed than lose consciousness. It took perseverance to get back to where he had been earlier. He was watching his legs toiled to make him stand. He looked at the window and look out but the narrow street was enveloped in tedious fog. He lays there quietly perhaps expected the total stillness to bring things back from his tedious state to natural one.

He was thinking if he falls out of bed all at the same time and kept his head raised he could avoid injuring it. Probably no loud voice will happen when falling onto the carpet. His main concern was for not to be rambunctious in this early time of the morning. Gregor has more perseverance to do the new method.

The door rang. "That'll be someone from work", he said to himself. For a moment, everything remained quiet. "They're not opening the door", Gregor said to himself, caught in some folly hope. Gregor hear the visitor's voice and he knew who it was -- the chief clerk himself. The chief clerk was rambunctious he knocked on the door very hard. Gregor fall from his bed.

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1 Gregor woke saw nightmares full of …….
a) stories
b) nonsense
c) bugs

2 Although Gregor has many legs, they are …….
a) weak
b) long
c) big

3 Gregor work in business that takes more…….. than other businesses
a time
b endurance
c money

4 Gregor boss's nature is ……
a) loud
b) peaceful
c) understanding

5 When Gregor looked out the window the weather was……
a) dull
b) nice
c) typical

6 Gregor have to……. to have social life because of his job.
a) plan
b) avoid
c) work hard

Table target word percentage, number of CQs and key answer of METAMORPHOSIS in session 12								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*veritably (adv)*	*actually*	6	*b*	*veritably*
				afflicted (adj)	*distressed*	*2*	*c*	*afflicted*
				cogent (adj)	*affective*	*3*	*b*	*cogent*
				sluggishly (adv)	*heavily*	*4*	*a*	*grievous*
				grievous (adj)	*serious*	*5*	*a*	*sluggishly*
				lucid (adj)	*obvious*	*6*	*c*	*lucid*

Metamorphosis

Session 12

Read the following text and carefully answer the questions.

"Something's fallen down in there", said the chief clerk. The chief clerk's firm footsteps could now be veritably heard in the next room. "Gregor", said his father "open the door, the chief clerk has come and wants a cogent reason why you are not at work. "He isn't well", said his mother to the chief clerk.
"I'll be there in a moment", said Gregor sluggishly.
"I hope it's nothing grievous. But on the other hand, I must say that if we people in commerce ever become slightly afflicted, we simply have to overcome it because of business considerations." Said the chief clerk ".

The chief clerk raised his voice, "Gregor! You are causing grievous concern to your parents and you fail to carry out your duties in a way that is veritably unheard of. I must request a Lucid explanation. This morning, your employer did suggest a cogent reason for your failure to appear. I intended to say all this to you in private, but since you cause me to waste my time here for no cogent reason I don't see why your parents should not also learn of it".
"But Sir", called Gregor, "I'm slightly afflicted and unwell, an attack of grievous sickness. I'm veritably fresh again now, though. I was veritably alright last night. I had a small symptom of it last night already. I don't know why I didn't let you know at work! But you always think you can get over being afflicted without staying at home".
And while Gregor gushed out these words, hardly knowing what he was saying. He wants to open the door. He was curious to learn what they would say when they caught sight of him. If they took everything calmly there is no cogent reason to be upset. Finally, he gave himself one last swing and stood there upright; the lower part of his body was in grievous pain.
"Oh, God!" called his mother, "he could be seriously afflicted and we're adding to his grievous suffer. Gregor is afflicted Ana, quick, get the doctor. Did you hear the way Gregor spoke just now?"
"That was the voice of an animal", said the chief clerk.
There was no lucid sound of the door banging shut again; they must have left it open.

Gregor had become much calm. So, they couldn't understand his words any more, although they seemed lucid enough to him, more lucid than before – perhaps his ears had become used to the sound. They had realised, though, that there was something wrong with him, and were ready to help.

Gregor sluggishly threw himself onto the door. Only holding himself upright with his mouth, and hung onto the key or pushed it down again with the whole weight of his body as needed. The lucid sound of the lock as it snapped back was Gregor's sign that he could veritably open the door Then he lay his head on the handle of the door to open it completely.

Because he had to open the door in this cogent way, it was already wide open before he could be seen. He had first to sluggishly turn himself around. He heard the chief clerk exclaim a loud "Oh!", he was the nearest to the door – his hand pressed against his open mouth and sluggishly retreating as if driven by a steady and invisible force. Gregor's mother, unfolded her arms, took two steps forward towards Gregor and sluggishly sank down onto the floor. His father looked hostile, and clenched his fists as if wanting to knock Gregor back into his room.

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1 When Gregor in his room he could ……heard the chief clerk's firm footsteps

a) almost
b) actually
c) not

2 The chief clerk said that even when people become slightly …….. they simply have to overcome it because of business considerations.

a) disrespected
b) board
c) distressed

3 The chief clerk wants ……. reason for Gregor failure to appear in the morning for his employer.

a) a real
b) an effective
c) a detailed

4 Gregor's mother was worried that he has been in …….. suffering,

a) serious
b) silent
c) full

5 Gregor…… threw himself onto the door.

a) heavily
b) slowly
c) immediately

6 The ……. Sound of the lock as it snapped back was Gregor's sign that he could open the door.

a) low
b) unclear
c) obvious

Session 13

Table target word percentage, number of CQs and key answer of METAMORPHOSIS in session 13								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
7	35	*700*	*7*	*veritably (adv)*	*actually*	*1*	*a*	*petrify*
				feeble (adj)	*weak*	*2*	*a*	*astound*
				petrify (v)	*frighten*	*3*	*a*	*expeditious*
				sluggishly (adv)	*heavily*	*4*	*a*	*veritably*
				astound (v)	*astonish*	*5*	*b*	*lucid*
				lucid (adj)	*obvious*	*6*	*c*	*feeble*
				expeditious (adj)	*speedy*	*7*	*c*	*sluggishly*

Metamorphosis

Session 13

Read the following text and carefully answer the questions.

But you, sir, you have a lucid overview than the rest of the staff, it is veritably a lucid overview than the boss himself – it's very easy for a businessman like him to make mistakes about his employees and make judgments more expeditious than you. Please, don't go away, at least first say something to show that you grant that I'm at least veritably right!"

But it seemed that I petrified him. My appearance had defiantly astound the man. The chief clerk had turned away as soon as Gregor had started to speak, and only stared back at him over his shaking shoulders as he left. He moved sluggishly towards the door without taking his eyes off him. He moved very gradually and sluggishly until he had reached the entrance hall and then he made an expeditious run, he rushed forward in a panic.

It was lucid to Gregor that he petrified the chief clerk and he decided to let him go. The chief clerk had to be held back, calmed down, convinced and finally won over; because it is lucid that the future of Gregor and his family depended on him! The chief clerk would listen to my sister, she is clever, he would not be more expeditious in his escape from Gregor.

Unfortunately, she wasn't there and Gregor sluggishly pushed himself through the door; tried to reach the chief clerk on the landing who veritably was holding on to the banister with both hands; but Gregor fell immediately over and landed on his numerous feeble little legs. He began to feel alright with his body; the feeble little legs had the solid ground under them. His mother was not far away in front of him and seemed that he astounds her, but then she was petrified by him and she shout Help, for pity's sake, Help!"

The look on his mother face-astound him and distract my attention. She was petrified by him and fled into the arms of his father who made an expeditious movement towards her. Gregor shouts "Mother! Please. Mother!" but he knows that he astounds his poor parents but he didn't have time to explain. The escape of the chief clerk petrified Gregor's father as well. His father picked up a large newspaper from the table with his left, and used them to drive Gregor back into his room. Gregor's appeals to his father were of no help, his appeals were simply not understood. Gregor realised that he was veritably incapable of going backwards in a straight line; he began, as quickly as possible to turn himself round. It went very sluggishly, but perhaps his father was able to see his good intentions. His father was hissing! It was making Gregor veritably confused. When he had nearly finished turning around with his feeble legs. His body was too broad to get through one of the double doors. In his present mood, it is lucid to him that it did not occur to his father to open the other of the double doors so that Gregor would have enough space to get through. Gregor pushed himself into the doorway without regard for what might happen to his feeble body. He falls in the ground, the feeble little legs along one side hung in the air while those on the other side pressed painfully against the ground. Then his father gave him a hefty shove from then the door was slammed shut. The father's acts astound Gregor and made him wonder how to get his father knows him.

II

Gregor awoke from his deep sleep. He pushed himself over to the door, feeling his way sluggishly with his antennae – of which he was now beginning to learn the value. He was even hungrier than he had been that morning, and immediately dipped his head into sweetened milk that was put in his room by his sister. But the milk did not taste at all nice. Later on, it seemed that

someone needed to enter the room but it is lucid the he thought he does not to be expeditious. The person who was trying to come in did not try to open the door again.
Now that you read the test please answer the following.

Circle the letter a, b, or c with the closest and most accurate completion of the sentence

1 Gregor seems to……. The chief clerk because he saw him staring back at him over his shaking shoulders as he left

a) frighten
b) please
c) follow

2 Gregor knows that he …… his poor parents but he didn't have time to explain.

a) shock
b) please
c) calm

3 The chief clerk would listen to my sister, she is clever, he would not be more ……. In his escape from Gregor.

a) speedy
b) slow
c) uncertain

4 Gregor father was hissing! It was making Gregor …….. confused.

a) actually
b) less
c) more

5 It was …… to Gregor that it did not occur to his father to open the other of the double doors so that he would have enough space to get through.

a) unclear
b) obvious
c) sad

6 Gregor pushed himself into the doorway without regard for what might happen to his …….. body.

a) big
b) strange
c) weak

7 Gregor …….. pushed himself through the door

a) easily
b) carefully
c) heavily

Session 14

Table target word percentage, number of CQs and key answer of METAMORPHOSIS in session 14								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
6	30	*600*	*6*	*toil (v)*	*work hard*	6	*c*	*tedious*
				tedious (adj)	*dull*	*2*	*b*	*expeditious*
				cogent (adj)	*affective*	*3*	*b*	*cogent*
				expeditious (adj)	*speedy*	*4*	*c*	*toil*
				grievous (adj)	*serious*	*5*	*a*	*astound*
				astound (v)	*astonish*	*6*	*c*	*grievous*

Metamorphosis

Session 14

Read the following text and carefully answer the questions.

It was clear that no one would come into Gregor's room until morning. That gave him time to think about this grievous change. For some reason, the empty room made him have tedious and uneasy feelings. Hardly aware of what he was doing, he made an expeditious movement and hurried under the couch. He felt immediately at ease.

He felt board as he spent the whole tedious night there. Some of the time sleeps although he frequently woke because of his hunger, and sometimes he spent in vague hopes which led to the same cogent conclusion: for the time being he must remain calm, he must show patience even if it is tedious to stay all this time by himself.

Next morning, Gregor's sister opened the door and when she noticed him under the couch, she slammed the door shut again from outside in expeditious way. But she opened the door again and came in on tip - toe as if entering the room of someone who has grievous illness. Gregor had toiled to push his head forward, right to the edge of the couch, and watched her. He felt an urge to make an expeditious rush forward at his sister's feet and beg her for something good to eat.

However, his sister astounded when she noticed that he did not drink the milk. She immediately picked the dish up and went out of the room. Gregor was extremely curious as to what she would bring in its place. He was astound when he saw his sister brought leftovers and rotten food. Then, out of cogent consideration for Gregor's feelings, she made an expeditious escape and locked the door. Gregor did not toil this time to move his body instead his little legs moved very easily, at last he could eat.

Gregor astounded when he found out that his injuries have completely healed. Quickly he consumed the leftovers and the rotten food. Then his sister slowly turned the key in the lock as a sign to him that he should withdraw. He found it cogent for his sister to make such gesture, and he hurried back under the couch. But he toiled to stay as his stomach is full food. He felt that he would suffer from grievous suffocation but he toiled to remain calm. His sister carried everything out leaving Gregor in his tedious room again.

This was how Gregor received his food each day now. Gregor's parents certainly did not want him to starve, but perhaps it would have been more cogent to be told about him than watching him. By time Gregor's sister got used to how he looks and when his brother sometimes does not finish his food, she would often say, sadly, "he must be bored from this tedious food". Whenever he heard anyone speaking in the next room he would made an expeditious walk straight to the door. For two whole days, all the talk at every mealtime was about the same grievous subject, me. There were always at least two members of the family. It was cogent that nobody wanted to be at home by themselves.

On the very first day the maid begged Gregor's mother to let her go without delay. Gregor astounded when she one day swore emphatically not to tell anyone the slightest about the grievous matter that happened. Now Gregor's sister also toil to help his mother with the cooking. On the first day, also he heard his father say something astounded him and it was a good new. He heard that his father's business can be strong again.\

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1 For some reason, the empty room made Gregor have …….. and uneasy feelings.

a) happy
b) unhealthy
c) dull

2 When Gregor's sister noticed him under the couch, she slammed the door shut again from outside in ……. way.

a) slow
b) fast
c) strange

3 Gregor's parents certainly did not want him to starve, but perhaps it would have been more …….. to be told about him than watching him.

a) serious
b) effective
c) disgusting

4 Gregor's sister……. to help his mother with the cooking

a) do not want
b) wish
c) work hard

5 Gregor ……. when he found out that his injuries have completely healed

a) astonished
b) frightened
c) explored

6 Gregor's sister came in the room on tip - toe as if entering the room of someone who has …….. illness

a) dangerous
b) kind of
c) serious

Table target word percentage, number of CQs and key answer of METAMORPHOSIS in session 15								
Number of target words	Repetition of target word 5%	Length of a text	CQs	target word	synonym	CQs number	Key answer	Indirect reference
5	25	*500*	*5*	*folly (n)*	*nonsense*	*1*	*b*	*perseverance*
				afflicted (adj)	*distressed*	*2*	*c*	*afflicted*
				perseverance (n)	*endurance*	*3*	*b*	*folly*
				rambunctious (adj)	*loud*	*4*	*a*	*petrify*
				petrify (v)	*frighten*	*5*	*b*	*rambunctious*

Metamorphosis

Session 15

Read the following text and carefully answer the questions.

Gregor worked with perseverance and converted his success at work straight into cash that delighted his family. The family got used to it and there was no longer much warm affection given to Gregor in return. Gregor did not think accomplishing his family dreams instead of his dreams is folly.

Gregor was concerned who will provide for his family. His father slow and clumsy, and lacking in perseverance. Gregor's mother cannot work. She is afflicted and suffered from asthma. Gregor's sister was still a child of seventeen, her life consisting luxury and joining in with folly activities. Whenever they began to talk of the need of money, Gregor became afflicted and quite hot with shame.

Alone in his room, Gregor tried not to produce any sound as his family become petrified and if they hear the slightest voice coming from the room. Gregor would go to all the effort try different moves without being rambunctious. He would surprise himself by letting go of the ceiling and landing on the floor with a rambunctious crash. But now, of course, he had far better control of his body

Gregor wanted to have more perseverance to live with his condition but it was hard to watch his sister caring for him. His observant sister leave the window open because she will become afflicted from his bad smell. She would alarm Gregor twice a day not to be rambunctious so no one scolds him. Gregor had left traces of the sticky substance from his feet as he crawled about his sister asked the maid as she will be petrified to be by herself in this folly matter. Gregor would love to thank his sister but not being able to speak to her made him afflicted.

By time Gregor's sister found it difficult to bare his smell. She is now rambunctious and she would clean and close the door fast and with loud noises. She seemed much more afflicted than usual. It made him realise that she is also still petrified by his look.

One day, about a month after Gregor's transformation when his sister feel that it was folly to be shocked at his appearance. His sister became less petrified to her brother’s look. His mother also wanted to go in and visit Gregor but his father petrified her at first and persuaded her against it. But then Gregor’s mother become rambunctious and shouted "Let me go and see Gregor, he is my unfortunate son!

Gregor liked his mother perseverance to see him. One day Gregor ‘s sister lead his mother in by the hand. Gregor hide under the sheets. The mother and sister thought it is better to move the old chest of drawers to make enough space for Gregor but it was too heavy. Both weak ladies pushed the chest with perseverance. After labouring at it for fifteen minutes his mother said it is folly to take the furniture away. However, this would have let him crawl around unrestricted in any direction.

Now that you read the test please answer the following.
Circle the letter a, b, or c with the closest and most accurate completion of the sentences

1 Gregor's father is slow and clumsy, and lacking in ……..

a) skills
b) determination
c) knowledge

2 Gregor's mother is ……. and suffered from asthma.

a) healthy
b) sad
c) distressed

3 Gregor's sister was still a child of seventeen, her life consisting luxury and joining in with ……. activities.

a) expensive
b) nonsenses
c) unnecessary

4 For a long time Gregor's sister still …….by his look

a) scared
b) worried
c) sad

5 Gregor would go to all the effort try different moves without being ……

a) hurt
b) loud
c) seen

Appendix 3. Material
D. Graded Readers' Target Words Distribution

WORD	texts														
	Free Culture First five days of learning					A Scandal in the Bohemia Second five days of learning					Metamorphosis Third five days of learning				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
vicissitude	5		5												
postulation	5				5										
contentious	5			5											
ingenious	5		5												
proprietorship	5		5												
concur		5	5												
new-fangled		5	5												
perceptive		5		5											
shrewd		5		5											
elucidate		5			5										
trespass		5		5											
contrive			5		5										
banish	5				5										
pertinent				5	5										
detestable				5	5										
adroit						5			5						
stupendous						5				5					
misgiving						5	5								
gargantuan						5			5						
irksome						5			5						
mystification						5	5								
ebullient						5		5							
idiosyncratic							5		5						
acrid							5		5						
obstinate							5	5							
colossal							5	5							
bonny								5		5					
apprehensive								5		5					
brittle								5		5					
unequivocally									5	5					
folly											5				5
feeble											5		5		
toil											5			5	
tedious											5			5	
perseverance											5				5
rambunctious											5				5
veritably												5	5		
grievous												5		5	
cogent												5		5	
afflicted												5			5
lucid												5	5		
sluggishly												5	5		
astound													5	5	
petrify													5		5
expeditious													5	5	

Appendix 3. Material
E. Graded Readers' recording sheet

WORD	Free Culture First five days of learning					A Scandal in the Bohemia Second five days of learning					Metamorphosis Third five days of learning				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
vicissitude															
postulation															
contentious															
ingenious															
proprietorship															
concur															
new-fangled															
perceptive															
shrewd															
elucidate															
trespass															
contrive															
banish															
pertinent															
detestable															
adroit															
stupendous															
misgiving															
gargantuan															
irksome															
mystification															
ebullient															
idiosyncratic															
acrid															
obstinate															
colossal															
bonny															
apprehensive															
brittle															
unequivocally															
folly															
feeble															
toil															
tedious															
perseverance															
rambunctious															
veritably															
grievous															
cogent															
afflicted															
lucid															
sluggishly															
astound															
petrify															
expeditious															

Appendix 4. Post-test
A. The LDT word and non-words order

1.

Pilot study LDT	
Trial name	**String type**
Instruction	
Vicissitude	Word
Proprietorship	Word
Postulations	Word
Galpin	Non-word
Moffat	Non-word
Help	Word
Contentious	Word
Ingenious	Word
Horozoone	Non-word
Venn	Non-word
Oestrogeny	Non-word
Beautiful	Word
Language	Word
New-fangled	Word
Perceptive	Word
Haque	Non-word
Lapidoscope	Non-word
Adair	Non-word
Welbbert	Non-word
Past	Word
Concur	Word
Elucidate	Word
Shrewd	Word
Trespass	Word
Opie	Non-word
Detailooring	Non-word
Acklon	Non-word
Stimulcrate	Non-word
Reservory	Non-word
Contrive	Word
Detestable	Word
Ralling	Non-word
Wray	Non-word
Berrow	Non-word
Healthy	Word
Woolnough	Non-word
Banish	Word

Pertinent	Word
Draconite	Non-word
Retrogradient	Non-word

2.

Main study LDT 1	
Trail name	**String type**
Instruction 1	
Instruction 2	
BANISH	Word
CONCUR	Word
VICISSITUDE	Word
DETESTABLE	Word
ACKLOON	Non-word
BODELATE	Non-word
HAPGOOD	Non-word
JUSTAL	Non-word
INGENIOUS	Word
CONTRIVE	Word
NEW-FANGLED	Word
PERTINENT	Word
VOLUNTEER	Word
TWOSE	Non-word
HOROZOONE	Non-word
GAPLINE	Non-word
COMPANY	Word
BEAUTIFUL	Word
LANGUAGE	Word
PROPRIETORSHIP	Word
PERCEPTIVE	Word
FLUCTUAL	Non-word
TRUGEON	Non-word
INTERNATIONAL	Word
BOOK	Word
ELUCIDATE	Word
TRESPASS	Word
CHAIR	Word
ECKETT	Non-word
MOFFAT	Non-word
SUDDERY	Non-word
OXYLATE	Non-word
SITUATION	word
POSTULATION	Word
CONTENTIOUS	Word
SHREWD	Word

BREAKFAST	Word
POTATO	Word
WOOLNOIUGH	Non-word
RECENTICLE	Non-word

3.

Main study LDT 2	
Trial name	**String type**
Instruction 1	
Instruction 2	
ACKLOON	Non-word
BODELATE	Non-word
BANISH	Word
LANGUAGE	Word
STUPENDOUS	Word
CAMBULE	Non-word
POCOCK	Non-word
HOROZOONE	Non-word
JUSTAL	Non-word
WOOLNOUGH	Non-word
ACRID	Word
IRKSOME	Word
VICISSITUDE	Word
CHAIR	Word
IDIOSYNCRATIC	Word
BONNY	Word
BOOK	Word
BUTTERFLY	Word
PARENT	Word
CONCUR	Word
GAPLIN	Non-word
CANTORAL	Non-word
DETESTABLE	Word
UNEQUIVOCALLY	Word
POSTULATION	Word
ZOO	Word
COLOSSAL	Word
BREAKFAST	Word
DINNER	Word
OBSTINATE	Word
ELUCIDATE	Word
MOFFAT	Non-word
SUDDERY	Non-word
TRUGEON	Non-word
TWOSE	Non-word

FLUCTUAL	Non-word
BEAUTIFUL	Word
KITCHEN	Word
RAINBOW	Word
NEW-FANGLED	Word
TRESPASS	Word
ANGRY	Word
ABROGATIVE	Non-word
MISGIVING	Word
PERCEPTIVE	Word
INGENIOUS	Word
SHREWD	Word
WILDING	Non-word
AISROPE	Non-word
VENN	Non-word
ALMANICAL	Non-word
WINTER	Word
MYSTIFICATION	Word
BRITTLE	Word
ADROIT	Word
GARGANTUAN	Word
RALLING	Non-word
STIMULCRATE	Non-word
BALDOCK	Non-word
CONTENTIOUS	Word
PROPRIETORSHIP	Word
EBULLIANT	Word
CANTILEEN	Non-word
OLIGATION	Non-word
PENCIL	Word
LETTER	word
TOOTH	word
SHIRT	Word
APPREHENSIVE	Word
CHARACTAL	Non-word
OESTROGENY	Non-word
DOGMATILE	Non-word
ESCROTAL	Non-word
OPIE	Non-word
RIDOUT	Non-word
PERTINENT	Word
INTERNATIONAL	Word
BANANA	Word
CONTRIVE	Word
SITUATION	Word

4.

Main Study LDT 3	
Trial name	**String type**
Instruction 1	
Instruction 2	
EBULLIENT	Word
CONTRIVE	Word
PERTINENT	Word
PROPRIETORSHIP	Word
SOFT	Word
RIDOUT	Non-word
VENN	Non-word
DESERET	Word
FOOD	Word
COGENT	Word
ASTOUND	Word
CAT	Word
DRESS	Word
BEDELATE	Non-word
ALDRED	Non-word
FLOSM	Non-word
TOIL	Word
AFFLICTED	Word
EXPEDITIOUS	Word
BANISH	Word
SWIM	Word
POCOCK	Non-word
HOROZOONE	Non-word
PROGRAM	Word
LANGUAGE	Word
STUPENDOUS	Word
RAMBUNCTIOUS	Word
FOLLY	Word
BAQULE	Non-word
GAMBASTYA	Non-word
JUSTAL	Non-word
COLD	Word
ACRID	Word
TABLE	Word
SLUGGISHLY	Word
GRIEVOUS	Word
RAFLOT	Non-word
PLUDOW	Non-word
CRATLY	Non-word

LINE	Word
IRKSOME	Word
VICISSITUDE	Word
CHAIR	Word
IDIOSYNCRATIC	Word
BOOVLE	Non-word
CHALT	Non-word
DAR	Non-word
HIMPOLA	Non-word
BONNY	Word
CLOCK	Word
PETRIFY	Word
INTERNATIONAL	Word
DEGATE	Non-word
MASAIL	Non-word
LIEL	Non-word
HAPGOOD	Non-word
PERSEVERANCE	Word
EDUCATION	Word
AFTERNOON	Word
CONCUR	Word
GAPLIN	Non-word
JABOLK	Non-word
OINIR	Non-word
BROWN	Word
DETESTABLE	Word
VERITABLY	Word
UNEQUIVOCALLY	Word
POSTULATION	Word
MOTHER	Word
TABOQ	Non-word
SARD	Non-word
DUFFIN	Non-word
COLOSSAL	Word
FEEBLE	Word
TEDIOUS	Word
BREAKFAST	Word
VOITLE	Non-word
TRAVE	Non-word
SIGENT	Non-word
ECKETT	Non-word
OBSTINATE	Word
ADROIT	Word
BRITTLE	Word
ELUCIDATE	Word

SUDDERY	Non-word
TRUGEON	Non-word
TWOSE	Non-word
GABOOSH	Non-word
NEW-FANGLED	Word
TRESPASS	Word
SHREWD	Word
BEAUTIFUL	Word
BLACKBOARD	Word
CARROT	Word
ALMANICAL	Non-word
OXYLATE	Non-word
ANGRY	Word
MISGIVING	Word
PERCEPTIVE	Word
INGENIOUS	Word
GLASS	Word
APRENE	Non-word
JARVIS	Non-word
STIMULCRATE	Non-word
WINTER	Word
MYSTIFICATION	Word
GARGANTUAN	Word
CONTENTIOUS	Word
CANTILEEN	Non-word
OLIGATION	Non-word
DETOOK	Non-word
TROAK	Non-word
PENCIL	Word
SITUATION	Word
APPREHENSIVE	Word
COMPANY	Word
VOLUNTEER	Word
LITERATURE	Word
LUCID	Word
TOAG	Non-word

5.

Main Study LDT 4	
Trial name	**String type**
Instruction 1	
Instruction 2	
CANTILEEN	Non-word
OLIGATION	Non-word
DETOOK	Non-word

TROAK	Non-word
PENCIL	Word
SITUATION	Word
APPREHENSIVE	Word
COMPANY	Word
VOLUNTEER	Word
LITERATURE	Word
PALPABLE	Word
TOAG	Non-word
ALMANICAL	Non-word
OXYLATE	Non-word
ANGRY	Word
MISGIVING	Word
PERCEPTIVE	Word
INGENIOUS	Word
GLASS	Word
APRENE	Non-word
JARVIS	Non-word
STIMULCRATE	Non-word
WINTER	Word
MYSTIFICATION	Word
GARGANTUAN	Word
CONTENTIOUS	Word
OBSTINATE	Word
ADROIT	Word
BRITTLE	Word
ELUCIDATE	Word
SUDDERY	Non-word
TRUGEON	Non-word
TWOSE	Non-word
GABOOSH	Non-word
NEW-FANGLED	Word
TRESPASS	Word
SHREWD	Word
BEAUTIFUL	Word
BLACKBOARD	Word
CARROT	Word
TABOQ	Non-word
SARD	Non-word
DUFFIN	Non-word
COLOSSAL	Word
FEEBLE	Word
TEDIOUS	Word
BREAKFAST	Word

VOITLE	Non-word
TRAVE	Non-word
SIGENT	Non-word
ECKETT	Non-word
DEGATE	Non-word
MASAIL	Non-word
LIEL	Non-word
HAPGOOD	Non-word
PERSEVERANCE	Word
EDUCATION	Word
AFTERNOON	Word
CONCUR	Word
GAPLIN	Non-word
JABOLK	Non-word
OINIR	Non-word
BROWN	Word
DETESTABLE	Word
VERITABLY	Word
UNEQUIVOCALLY	Word
POSTULATION	Word
MOTHER	Word
RAFLOT	Non-word
PLUDOW	Non-word
CRATLY	Non-word
LINE	Word
IRKSOME	Word
VICISSITUDE	Word
CHAIR	Word
IDIOSYNCRATIC	Word
BOOVLE	Non-word
CHALT	Non-word
DAR	Non-word
HIMPOLA	Non-word
BONNY	Word
CLOCK	Word
PETRIFY	Word
INTERNATIONAL	Word
PROGRAM	Word
LANGUAGE	Word
STUPENDOUS	Word
RAMBUNCTIOUS	Word
FOLLY	Word
BAQULE	Non-word
GAMBASTYA	Non-word
JUSTAL	Non-word

COLD	Word
ACRID	Word
TABLE	Word
SLUGGISHLY	Word
GRIEVOUS	Word
EBULLIENT	Word
CONTRIVE	Word
PERTINENT	Word
PROPRIETORSHIP	Word
SOFT	Word
RIDOUT	Non-word
VENN	Non-word
DESERET	Word
FOOD	Word
COGENT	Word
ASTOUND	Word
CAT	Word
DRESS	Word
BEDELATE	Non-word
ALDRED	Non-word
FLOSM	Non-word
TOIL	Word
AFFLICTED	Word
EXPEDITIOUS	Word
BANISH	Word
SWIM	Word
POCOCK	Non-word
HOROZOONE	Non-word

Appendix 4. Post-test
B. LDT number of syllables in stimuli

1.

pilot LDT number of syllables in stimuli

number of syllables	**5 syllables**	**4 syllables**	**3 syllables**	**2 syllables**	**1 syllable**
number of target words	1	4	5	4	1
target words	pro.pri.e.tor.ship	vi.cis.si.tude pos.tu.la.tions de.test.a.ble e.lu.ci.date	con.ten.tious new.fan.gled per.cep.tive in.gen.ious per.ti.nent	con.cur con.trive ban.ish tres.pass	shrewd
number of words			1	2	2
words			beau.ti.ful	health.y lan.guage	help past
number of non-words		1	6	8	5
non-word		de.tail.o.oring 4	oest.ro.geny lapid.o.scope stim.ul.crate res.er.vory dra.con.ite retro.grad.ient	horo.zoone welb.bert gap.lin ack.lon ral.ling ber.row wool.nough mof.fat	venn haque adair opie wray

2.

main study LDT 1 number of syllables in stimuli

number of syllables	5 syllables	4 syllables	3 syllables	2 syllables	1 syllable
number of target words	1	4	5	4	1
target words	pro.pri.e.tor.ship	vi.cis.si.tude pos.tu.la.tions de.test.a.ble e.lu.ci.date	con.ten.tious new.fan.gled per.cep.tive in.gen.ious per.ti.nent	con.cur con.trive ban.ish tres.pass	shrewd
number of words			1	2	2
words	in.ter.na.tion.al	sit.u.a.tion	beau.ti.ful po.ta.to com.pa.ny vol.un.teer	lan.guage break.fast	book chair
number of non-words			2	17	1
non-word			re.cent.icle bo.de.late	horo.zoone gap.lin wool.nough ack.loon hap.good fluc.tual sud.dery oxy.late trug.eon just.al e.ckett	twose

main study LDT 1 number of syllables in stimuli

number of syllables	5 syllables	4 syllables	3 syllables	2 syllables	1 syllable
				mof.fat	

2.

main study LDT 2 number of syllables in stimuli

number of syllables	6 syllables	5 syllables	4 syllables	3 syllables	2 syllables	1 syllable
number of target words	2	2	5	11	9	1
target words	id.i.o.syn.crat.ic un.e.quiv.o.cal.ly	pro.pri.e.tor.ship mys.ti.fi.ca.tion	vi.cis.si.tude pos.tu.la.tion de.test.a.ble e.lu.ci.date ap.pre.hen.sive	con.ten.tious new.fan.gled per.cep.tive in.gen.ious per.ti.nent stu.pen.dous co.los.sal ob.sti.nate gar.gan.tuan e.bul.lient mis.giv.ing	con.cur con.trive ban.ish tres.pass ac.rid irk.some bon.ny brit.tle ad.roit	shrewd
number of words		1	1	3	10	5
words		in.ter.na.tion.al	sit.u.a.tion	beau.ti.ful but.ter.fly ba.na.na	lan.guage break.fast kit.chen rain.bow win.ter par.ent din.ner pen.cil let.ter	book chair zoo tooth shirt

main study LDT 2 number of syllables in stimuli

number of syllables	6 syllables	5 syllables	4 syllables	3 syllables	2 syllables	1 syllable
					an.gry	
number of non-words			1	9	17	3
non-word			al.ma.ni.cal	oest.ro.geny stim.ul.crate abro.ga.tive oli.ga.tion dog.ma.tile es.cro.tal can.to.ral fluc.tu.al can.ti.leen	mof.fat horo.zoone gap.lin ack.loon ral.ling wool.nough bode.late cam.bule sud.dery poc.ock trug.eon wild.ing ais.rope bald.ock chara.ctal rid.out jus.tal	venn opie twose

3.

main study LDT 3 and LDT 4 number of syllables in stimuli

number of syllables	6 syllables	5 syllables	4 syllables	3 syllables	2 syllables	1 syllable
number of words	2	2	9	16	14	2
target words	id.i.o.syn.crat.ic un.e.quiv.o.cal.ly	pro.pri.e.tor.ship mys.ti.fi.ca.tion	vi.cis.si.tude pos.tu.la.tions de.test.a.ble	con.ten.tious new.fan.gled per.cep.tive	con.cur con.trive ban.ish	shrewd toil

main study LDT 3 and LDT 4 number of syllables in stimuli

number of syllables	6 syllables	5 syllables	4 syllables	3 syllables	2 syllables	1 syllable
			e.lu.ci.date	in.gen.ious	tres.pass	
			ap.pre.hen.sive	per.ti.nent	as.tound	
			ex.pe.di.tious	a.dr.oit 2	ac.rid	
			per.se.ver.ance	stu.pen.dous	irk.some	
			ver.i.ta.bly	co.los.sal	co.gent	
			gar.gan.tu.an	ob.sti.nate	griev.ous	
				e.bul.lient	bon.ny	
				mis.giv.ing	brit.tle	
				af.flict.ed	fol.ly	
				ram.bunc.tious	feeb.le	
				slug.gish.ly	lu.cid	
				pet.ri.fy		
				te.di.ous		
number of words			2	2	10	16
words		in.ter.na.tion.al	ed.u.ca.tion	beau.ti.ful	moth.er	chair
			lit.er.a.ture	af.ter.noon	lan.guage	glass
			sit.u.a.tion	com.pa.ny	break.fast	soft
				vol.un.teer	win.ter	food
					de.sert	cat
					pro.gram	dress
					an.gry	swim
					pen.cil	cold
					ta.ble	clock
					balck.board	carrot
						line
						brown
number of non-words			1	5	30	9

main study LDT 3 and LDT 4 number of syllables in stimuli

number of syllables	6 syllables	5 syllables	4 syllables	3 syllables	2 syllables	1 syllable
non-word			o.li.ga.tion	e.ck.ett	horo.zoone	toag
				stim.ul.crate	voit.le	venn
				al.man.ical	ap.rene	sard
				gam.ba.stya	jar.vis	twose
				be.del.ate	gap.lin	duffin
					tab.oq	chalt
					hap.good	dar
					de.gate	trave
					oxy.late	troak
					poc.ock	
					trug.eon	
					canti.leen	
					rid.out	
					ald.red	
					ba.qule	
					plu.dow	
					boo.vle	
					him.pola	
					flo.sm	
					raf.lolt	
					crat.ly	
					jus.tal	
					ma.sail	
					li.el	
					jab.olk	
					sig.ent	
					oin.ir	
					sud.dery	
					ga.boosh	
					de.took	

Appendix 4. Post-test
C. The Gap-fill Tests

1. Pilot GFT and main study GFT 1

Name:……..

Complete the following sentences with a suitable word. Every blank provide the first three or four letters of the target word as a hint for you.

1. Some people enjoy the vici…… of seasons.
2. Propo……. is a type of business entity that is owned and run by a person.
3. I do not agree with the post……. of this book.
4. There will always be cont…….. issues regarding religion and faith.
5. He was inge…… enough to overcome his limitations.
6. I've no time for such new-f…….. celebrations.
7. Gada Alsman is a perce…… writer.
8. The new results conc…….. with the findings from previous research.
9. I don't understand you will have to eluc…….. . I need to know more details .
10. Alwaleed bin Talal is a shr…… businessman.
11. The government will not allow enemy to tres…….. to our land.
12. The author cont…….. a clever plot.
13. Humanity should stand against dete…… wars.
14. Some people were ban…….. to Siberia for political crimes.
15. She asks the bride many perti…….. details.

2.

Main study GFT 2

Name:……..

Complete the following sentences with a suitable word. Every blank provide the first three or four letters of the target word as a hint for you.

1. Some people enjoy the vici…… of seasons.
2. Propo……. is a type of business entity that is owned and run by a person.
3. I do not agree with the post……. of this book.
4. There will always be cont…….. issues regarding religion and faith.
5. He was inge…… enough to overcome his limitations.
6. I've no time for such new-f…….. celebrations.
7. Gada Alsman is a perce…… writer.
8. The new results conc…….. with the findings from previous research.
9. I don't understand you will have to eluc…….. . I need to know more details .
10. Alwaleed bin Talal is a shr…… businessman.
11. The government will not allow enemy to tres…….. to our land.
12. The author cont…….. a clever plot.
13. Humanity should stand against dete…… wars.
14. Some people were ban…….. to Siberia for political crimes.
15. She asks the bride many perti…….. details.
16. My Lawyer is an adro…….. debater.
17. Founding everyone alive after the disaster is a stupe……. news.
18. When I learned my daughter's fiancé is arrested for careless driving, my misg…… about him became stronger.
19. It took three men to carry the garga……. couch into the house.
20. Many irks…… noise is coming from the building next to our house.
21. The Arab spring is one of the great mystif…… in our history.
22. Mnsour always felt ebul…… after drinking his afternoon tea.
23. An example of an idios……. person is someone who does things different than the norm.
24. The acr…….. smell of the rotten food is enough to make anyone vomit.
25. He showed an obst……. deny to admit that it is his fault.
26. The colo……. eruption of the volcano, mad everyone scream.
27. I saw a bon……. child dressed in white.
28. When he learned about the police arrest, his face expression was appre…….
29. The old lady's bone became fragile and bri……
30. We uneq…… condemn any child abuse.

3.

Main study GFT 3 and GFT 4

Name ……

Complete the following sentences with a suitable word. Every blank provide the first three or four letters of the target word as a hint for you.

1. Some people enjoy the vici…… of seasons.
2. Propo……. is a type of business entity that is owned and run by a person.
3. I do not agree with the post……. of this book.
4. There will always be cont…….. issues regarding religion and faith.
5. He was inge…… enough to overcome his limitations.
6. I've no time for such new-f…….. celebrations.
7. Gada Alsman is a perce…… writer.
8. The new results conc…….. with the findings from previous research.
9. I don't understand you will have to eluc…….. . I need to know more details .
10. Alwaleed bin Talal is a shr…… businessman.
11. The government will not allow enemy to tres…….. to our land.
12. The author cont…….. a clever plot.
13. Humanity should stand against dete…… wars.
14. Some people were ban…….. to Siberia for political crimes.
15. She asks the bride many perti…….. details.
16. My Lawyer is an adro…….. debater.
17. Founding everyone alive after the disaster is a stupe……. news.
18. When I learned my daughter's fiancé is arrested for careless driving, my misg…… about him became stronger.
19. It took three men to carry the garga……. couch into the house.
20. Many irks…… noise is coming from the building next to our house.
21. The Arab spring is one of the great mystif…… in our history.
22. Mnsour always felt ebul…… after drinking his afternoon tea.
23. An example of an idios……. person is someone who does things different than the norm.
24. The acr…….. smell of the rotten food is enough to make anyone vomit.
25. He showed an obst……. deny to admit that it is his fault.
26. The colo……. eruption of the volcano, mad everyone scream.
27. I saw a bon……. child dressed in white.
28. When he learned about the police arrest, his face expression was appre…….
29. The old lady's bone became fragile and bri……
30. We uneq…… condemn any child abuse.
31. After the accident, Khaled had suffered a grie…….wound that he would take a long time to be treated in the hospital.
32. There is a cog……. argument to be made that the price of cigarette should increase.
33. Ahmad had to to…….. for many years before becoming a famous doctor.
34. I did not enjoy this tedi……. journey.
35. It is not fare to make this expe……..investigation in such an important accusation.

36. The taste of the delicious cooking is going to asto…….. anyone who likes fine dining.
37. Succeeding in high school is verit…….. what she is working hard for.
38. The old lady was too fee…… to get out of her bed.
39. His younger son was affli…….. with a serious sickness.
40. The fat man slug……. moved himself from the couch to the bed.
41. You shouldn't pet……. poor little Fatimah with your silly prank.
42. She gave a luc…… presentation to her colleagues.
43. She made herself proud through hard work and perse.....
44. It would be fol…….. to spend money carelessly.
45. It is normal to have ramb……. kids in Eid celebrations.

Appendix 5. Pilot Study Results
A. RT outliers

1.

Mean and standard deviations of RT with outliers.			
Group	Word card groups		GR
	RM	FD	
Mean	3104.38	3995.36	3288.23
SD	1264.70	1512.82	1316.41

2.

Mean and standard deviations of reaction time **without** outliers.			
Group	Word card groups		GR
	RM	FD	
Mean	3104.38	3354.50	2938.51
SD	1264.70	762.494	757.40

- FD: 6335.80 and 6781.87 were replaced by 3354.50
- GR: 6436 was replaced by 2938.51

Appendix 5. Pilot Study Results
B. FWMs and BWMs for the RMCG and the FDCG

1.

Means and standard deviation of the FWMs for the RMCG and the FDCG in the pilot study.

word	RMCG		FDCG	
	Mean	SD	Mean	SD
vicissitude	5.1	0.699	5.1	0.316
proprietorship	5.6	0.843	5.4	0.843
postulation	4.6	0.568	5.4	0.699
contentious	5.1	0.000	5	0.471
ingenious	5	0.422	5	0.000
new-fangled	5.2	0.516	5.2	0.632
perceptive	5.4	0.316	5	0.471
concur	4.9	0.667	5.1	0.316
elucidate	5	0.000	5.2	0.422
shrewd	5	0.422	5	0.000
trespass	5.2	0.316	5.1	0.316
contrive	5.1	0.483	5	0.000
detestable	5.3	0.316	5.2	0.632
banish	5.1	0.789	5.4	0.699
pertinent	5.2	0.699	4.8	0.422

2.

Means and standard deviation of the BWMs for RMCG and FDCG in the pilot study

word	RMCG		FDCG	
	Mean	SD	Mean	SD
vicissitude	0.30	0.483	0.60	0.843
proprietorship	0.60	0.699	0.60	0.699
postulation	0.20	0.422	0.70	0.675
contentious	0.50	0.527	0.30	0.483
ingenious	0.00	0.000	0.00	0.000
new-fangled	0.20	0.422	0.20	0.632
perceptive	0.40	0.516	0.20	0.422
concur	0.00	0.000	0.10	0.316
elucidate	0.40	0.699	0.20	0.422
shrewd	0.00	0.000	0.00	0.000
trespass	0.20	0.422	0.10	0.316

contrive	0.20	0.632	0.10	0.316
detestable	0.40	0.516	0.30	0.675
banish	0.10	0.316	0.50	0.972
pertinent	0.40	0.516	0.30	0.483

Appendix 6. Main Study Results
A. RT outliers

1. LDT 1 RT outliers

RMCG LDT 1 RT outliers and new means (ms)

word	Outlier	New mean (replaced outlier)
vicissitude	4120	2279.67
proprietorship	4800	1909.75
postulation	5677	3158.28
	4045	2322.68
contentious	4126	2322.68
new-fangled	4909	2440.58
perceptive	5908	2621.43
concur	3603	2054.14
elucidate	5678	2419.79
shrewd	3489	1862.39
	3907	1882.14
trespass	4150	1882.14
detestable	5435	3273.17
pertinent	5090	2656.94

GRG LDT 1 RT outliers and new means (ms)

word	Outlier	New mean (replaced outlier)
vicissitude	4777	2699.94
proprietorship	6012	2281.37
new-fangled	4087	2508.52
perceptive	5570	3048.58
concur	3902	2287.10
elucidate	6545	3659.95
trespass	8643	3252.29
detestable	6534	3235.53
	6541	3235.53
banish	5226	2764.59
pertinent	10020	2771.83

CG LDT 1 RT outliers and new means (ms)

word	Outlier	New mean (replaced outlier)
concur	4945	2129.38
detestable	5404	2621.38

1. LDT 2 RT outliers

RMCG LDT 2 RT outliers and new means (ms)

word	outliers	mean (replaced outlier)
proprietorship	4892	2321.50
contentious	5403	2154.65
new-fangled	4789	2262.76
perceptive	16286	3342.55
concur	6278	1830.10
elucidate	5101	2220.67
contrive	4680	2212.29
detestable	6302	2750.90
pertinent	4511	2227.05
adroit	4568	2304.00
stupendous	4309	2174.79
gargantuan	4830	2214.52
irksome	4892	1708.95
ebullient	5397	2225.47
obstinate	4390	1958.29
colossal	5902	2141.68
bonny	3819	1788.00
	3791	1788.00
apprehensive	6987	2611.80
brittle	5873	2283.85

GRG LDT 2 RT outliers and new means (ms)

word	outliers	mean (replaced outlier)
vicissitude	3920	2082.00
ingenious	4390	2160.00
new-fangled	894	2262.76
	800	2262.76
perceptive	5029	2608.80
concur	4200	2042.29
elucidate	5920	3062.76
	6001	3062.76
shrewd	3744	1605.40
	3901	1605.40
pertinent	11591	3423.54
misgiving	5390	2274.92
gargantuan	5480	2371.05

irksome	5930	2062.59
mystification	6538	2117.64
ebullient	5893	1899.45
acrid	5794	2670.88
obstinate	4729	2168.42
colossal	5849	2111.53
	6940	2111.53
bonny	5351	1962.37
apprehensive	6191	2926.79
brittle	6782	2808.40
	6789	2808.40

CG LDT 2 RT outliers and new means (ms)

word	outliers	mean (replaced outlier)
proprietorship	7160	3073.91
postulation	5940	2521.75
contentious	5091	1946.00
perceptive	4290	2169.44
concur	4945	1465.00
shrewd	4602	2152.70
banish	3573	1240.45
adroit	5789	2474.50
misgiving	4892	2222.63
irksome	4392	1964.56
obstinate	3289	1538.36
colossal	5429	2160.00
bonny	3462	1624.91
apprehensive	7181	3021.71
brittle	6209	2295.25
unequivocally	6302	1915.45

2. LDT 3 RT outliers

RMCG LDT 3 RT outliers and new means replaced outlier (ms)

word	outliers	mean (replaced outlier)
proprietorship	3820	1718.59
postulation	3970	1703.36
contentious	6904	1810.41
new-fangled	5190	1812.58
	5050	
perceptive	5901	2093.62
concur	3190	1713.68

contrive	3401	1614.19
detestable	3892	2016.15
banish	2910	1392.73
pertinent	3299	1801.15
adroit	4930	2231.85
misgiving	4006	2211.00
mystification	4737	2155.09
ebullient	6782	2508.82
idiosyncratic	5176	2523.67
acrid	3292	1744.04
obstinate	3920	2070.13
bonny	3921	1723.33
unequivocally	4903	2140.40
	4983	
grievous	5665	2394.71
veritably	5147	2372.56
feeble	5008	1970.06
afflicted	14449	3461.37
sluggishly	9878	2954.91
	8229	
lucid	7860	2372.94
perseverance	5443	2061.80
folly	9385	2443.52
rambunctious	8076	2696.47

GRG LDT 3 RT outliers and new means (ms)

word	**outliers**	**mean (replaced outlier)**
vicissitude	3890	1993.437
	3920	
postulation	5748	2101.670
contentious	5489	2090.060
	5011	
ingenious	4390	1713.045
	5490	
perceptive	7009	2813.187
concur	4092	1879.610
shrewd	3290	1571.700
adroit	5360	2983.812
stupendous	5920	2714.070
gargantuan	7444	3427.000
irksome	4393	2298.560

mystification	6283	2629.820
ebullient	4362	2338.300
idiosyncratic	6271	2802.140
acrid	4269	1956.330
obstinate	9712	2719.720
colossal	4863	2246.680
apprehensive	5007	2254.530
brittle	6630	2978.740
unequivocally	8959	3902.540
grievous	15093	3548.060
cogent	5748	2408.290
toil	6755	2593.800
tedious	8540	2921.450
expeditious	6414	2884.170
	5899	
astound	4828	2342.570
veritably	6564	2617.000
feeble	4324	2248.200
afflicted	5405	2034.690
sluggishly	7804	2740.900
petrify	7085	2705.190
lucid	9562	2421.850
folly	10088	3068.050
rambunctious	19383	4250.380

CG LDT 3 RT outliers and new means (ms)

word	**outliers**	**mean (replaced outlier)**
proprietorship	2793	1488.42
contentious	11591	2600.44
shrewd	4539	1886.42
detestable	7192	2435.00
stupendous	3278	1144.00
misgiving	4839	1848.58
mystification	5903	2228.64
ebullient	10022	3107.13
obstinate	9301	2867.00
colossal	7647	2376.30
bonny	7381	2230.00
brittle	2331	1401.56
unequivocally	4832	2197.45
grievous	3829	1904.30

toil	4251	1941.50
lucid	3672	1575.67
perseverance	4781	1893.75
rambunctious	12737	3028.38

3. Post-test LDT 4 RT outliers

RMCG LDT4 RT outlier and new means (ms)		
words	**outlier**	**mean**
postulation	6234	2478.100
contentious	6555	2125.400
ingenious	3498	1591.450
new-fangled	3751	1671.250
	3533	
perceptive	4467	2013.750
concur	3798	1737.421
elucidate	8192	2414.150
shrewd	3906	2057.909
trespass	5409	1670.333
	5901	
detestable	7399	2149.315
pertinent	7089	2339.650
	6994	
adroit	6343	1637.470
stupendous	4891	1942.263
misgiving	4777	1661.470
mystification	3115	1367.823
ebullient	5998	1818.222
idiosyncratic	2330	1295.850
acrid	2710	1596.200
obstinate	3015	1638.052
colossal	7109	1727.600
grievous	4462	1676.200
cogent	5701	1955.461
afflicted	3290	1607.230
	3291	
sluggishly	7067	2197.684
petrify	4582	1817.470
lucid	3903	2556.722
perseverance	9991	2046.062
rambunctious	6090	2122.235

GRG LDT 4 RT outliers and new means (ms)

words	outlier	mean
proprietorship	4955	1975.056
postulation	4028	1930.750
contentious	4044	1393.571
ingenious	3845	1360.250
	3593	
new-fangled	3571	1361.714
perceptive	4219	1888.941
concur	3490	1668.375
elucidate	3991	1737.706
shrewd	2710	1395.294
trespass	4771	1902.778
detestable	5461	2377.688
banish	3587	1507.625
adroit	4072	1363.313
misgiving	7158	2148.688
gargantuan	5286	2469.889
irksome	4792	2132.600
mystification	4512	2018.188
ebullient	3584	1723.071
idiosyncratic	3300	1659.941
acrid	2965	1444.833
obstinate	4309	1464.143
	4836	
colossal	4440	2035.375
apprehensive	5215	2266.375
brittle	2819	1281.647
unequivocally	4587	1548.875
grievous	4568	2073.933
cogent	4248	1713.600
toil	4160	1825.529
tedious	3988	1828.056
sluggishly	5490	1818.647
lucid	2560	1280.833
	2523	
perseverance	3856	1679.500
rambunctious	3877	1862.733

CG LDT 4 RT outliers and new means (ms)

words	outlier	mean
vicissitude	2581	1220.692
postulation	3897	1543.600
perceptive	4900	1711.000
concur	3756	1418.444
shrewd	4245	1593.176
trespass	3769	1480.667
contrive	3091	1311.333
detestable	2910	1276.000
banish	4090	1290.428
pertinent	3891	1470.667
irksome	2956	1475.583
acrid	4917	1326.333
obstinate	5348	1636.500
colossal	7403	2235.778
apprehensive	5458	1443.444
brittle	6609	2257.800
unequivocally	3910	1398.300
grievous	5204	1678.909
toil	3595	1516.000
veritably	3789	1644.357
perseverance	4999	1773.272

Appendix 6. Main Study Results
B. RMCG FWMs and BWMs in each week

1. FWMs in week 1

Frequency of FWMs in the first week of learning.

Target words	**Number of FWMs**								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**mean**
vicissitude				1	20	4			5.12
proprietorship				1	14	7	2	1	5.52
postulation			1	4	15	4	1		5.00
contentious				2	20	3			5.04
ingenious					25				5.00
new-fangled					20	4	1		5.24
perceptive				1	19	5			5.16
concur				2	22	1			4.96
elucidate				2	19	4			5.08
shrewd					25				5.00
trespass					22	3			5.12
contrive					24	1			5.04
detestable				1	19	4	1		5.04
banish					19	4	1	1	5.36
pertinent				4	15	6			5.08

2. BWMs in week 1

Frequency of BWMs in the first week of learning.

Target words	**Number of BWMs**				
	no move	**1 moves**	**2 moves**	**3 moves**	**mean**
vicissitude	17	6	2		0.40
proprietorship	13	8	3	1	0.68
postulation	16	8	1		0.40
contentious	17	8			0.32
ingenious	25				0.00
new-fangled	20	4	1		0.24
perceptive	19	6			0.24
concur	23	2			0.08
elucidate	20	4	1		0.24
shrewd	25				0.00
trespass	22	3			0.12
contrive	23	1	1		0.12

detestable	17	7	1		0.36
banish	19	4		2	0.40
pertinent	15	9	1		0.44

3. FWMs in week 2

Frequency of FWMs in the second week of learning.

Target words	**Number of FWMs**								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	mean
vicissitude	2	1							0.16
proprietorship	1	1							0.12
postulation	4	1							0.24
contentious	2								0.08
ingenious									0.00
new-fangled									0.00
perceptive	2								0.08
concur	1		1						0.16
elucidate	3								0.12
shrewd									0.00
trespass									0.00
contrive	1								0.04
detestable	1								0.04
banish									0.00
pertinent	3	2							0.28
adroit				2	18	4	1		5.16
stupendous					21	2	2		5.24
misgiving					23	1	1		5.12
gargantuan					17	5	3		5.44
irksome					21	3	1		5.20
mystification				2	20	2	1		5.08
ebullient					21	2	2		5.24
idiosyncratic				4	15	6			5.08
acrid					21	3	1		5.20
obstinate					24		1		5.08
colossal					24		1		5.08
bonny					24	1			5.04
apprehensive					20	4	1		5.24
brittle					23	1	1		5.12
unequivocally					23	2			5.08

4. BWMs in week 2

Frequency of BWMs in the second week of learning.

Target words	Number of BWMs				
	no move	1 moves	2 moves	3 moves	mean
vicissitude	25				0.00
proprietorship	25				0.00
postulation	25				0.00
contentious	25				0.00
ingenious	25				0.00
new-fangled	25				0.00
perceptive	25				0.00
concur	25				0.00
elucidate	25				0.00
shrewd	25				0.00
trespass	25				0.00
contrive	25				0.00
detestable	25				0.00
banish	25				0.00
pertinent	25				0.00
adroit	20	4	1		0.25
stupendous	21	3	1		0.20
misgiving	21	2	2		0.24
gargantuan	17	6	2		0.40
irksome	21	4			0.16
mystification	21	3	1		0.20
ebullient	19	3	3		0.36
idiosyncratic	19	6			0.24
acrid	21	3	1		0.20
obstinate	23	1	1		0.12
colossal	24	1			0.04
bonny	24	1			0.04
apprehensive	20	4	1		0.24
brittle	22	2	1		0.16
unequivocally	23	2			0.08

5. FWMs in week 3

Frequency of FWMs in the third week of learning.

Target words	Number of repetition								
	1	2	3	4	5	6	7	8	mean
adroit	2								0.08
stupendous									0.00
misgiving	1								0.04
gargantuan									0.00
irksome	1								0.04
mystification	4								0.16
ebullient	2								0.08

idiosyncratic	2								0.08
acrid									0.00
obstinate									0.00
colossal									0.00
bonny									0.00
apprehensive	1								0.04
brittle	1								0.04
unequivocally									0.00
grievous			1		19	5			5.12
cogent				1	18	6			5.20
toil					19	5	1		5.28
tedious					22	3			5.12
expeditious					23	2			5.08
astound					20	3	2		5.28
veritably					21	3		1	5.24
feeble					23	2			5.08
afflicted					21	4			5.16
sluggishly					22	3			5.12
petrify					22	3			5.12
lucid					23	2			5.08
perseverance				2	20	3			5.04
folly					23	2			5.08
rambunctious					21	3	1		5.20

6. BWMs in week 3

Frequency of BWMs in the third week of learning.

Target words	**Number of backwards moves**				
	no move	1 moves	2 moves	3 moves	mean
grievous	18	7			0.28
cogent	18	7			0.28
toil	17	7	1		0.36
tedious	22	3			0.12
expeditious	22	3			0.12
astound	21	2	2		0.24
veritably	21	3		1	0.24
feeble	23	2			0.08
afflicted	20	5			0.20
sluggishly	22	3			0.12
petrify	22	3			0.12
lucid	22	3			0.12
perseverance	18	7			0.28
folly	23	2			0.08
rambunctious	20	4	1		0.24

Appendix 6. Main Study Results

C. Total frequency of FWMs and the BWMs by the end of the learning period

1. Total and means of FWMs in each target word at the end of the three weeks

RMCG total and means of FWMs in the three word lists in the three weeks of learning

List A	FW (total)	FW (mean)	List B	FW (total)	FW (mean)	List C	FW (total)	FW (mean)
vicissitude	133	5.32	*adroit*	131	5.24	*grievous*	128	5.12
proprietorship	141	5.64	*stupendous*	131	5.24	*cogent*	130	5.20
postulation	131	5.24	*misgiving*	129	5.16	*toil*	132	5.28
contentious	129	5.16	*gargantuan*	136	5.44	*tedious*	128	5.12
ingenious	125	5.00	*irksome*	131	5.24	*expeditious*	127	5.08
new-fangled	131	5.24	*mystification*	131	5.24	*astound*	132	5.28
perceptive	131	5.24	*ebullient*	133	5.32	*veritably*	131	5.24
concur	128	5.12	*idiosyncratic*	132	5.28	*feeble*	127	5.08
elucidate	130	5.20	*acrid*	130	5.20	*afflicted*	129	5.16
shrewd	125	5.00	*obstinate*	127	5.08	*sluggishly*	128	5.12
trespass	128	5.12	*colossal*	127	5.08	*petrify*	128	5.12
contrive	127	5.08	*bonny*	126	5.04	*lucid*	127	5.08
detestable	131	5.24	*apprehensive*	132	5.28	*perseverance*	126	5.04
banish	134	5.36	*brittle*	129	5.16	*folly*	127	5.08
pertinent	134	5.36	*unequivocally*	127	5.08	*rambunctious*	130	5.20

2. Total and means of BWMs in each target word at the end of the three weeks

RMCG total and means of BWMs in the three word lists in the three weeks of learning

List A	BW (total)	BW (mean)	List B	BW (total)	BW (mean)	List C	BW (total)	BW (mean)
vicissitude	10	.40	*adroit*	6	.24	*grievous*	7	.28
proprietorship	17	.68	*stupendous*	5	.20	*cogent*	7	.28

postulation	10	.40	*misgiving*	5	.20	*toil*	9	.36
contentious	8	.32	*gargantuan*	10	.4	*tedious*	3	.12
ingenious	0	.00	*irksome*	4	.16	*expeditious*	3	.12
new-fangled	6	.24	*mystification*	5	.20	*astound*	6	.24
perceptive	7	.28	*ebullient*	9	.36	*veritably*	6	.24
concur	2	.08	*idiosyncratic*	6	.24	*feeble*	2	.08
elucidate	6	.24	*acrid*	5	.20	*afflicted*	5	.20
shrewd	0	.00	*obstinate*	3	.12	*sluggishly*	3	.12
trespass	3	.12	*colossal*	2	.08	*petrify*	3	.12
contrive	3	.12	*bonny*	1	.04	*lucid*	3	.12
detestable	9	.36	*apprehensive*	6	.24	*perseverance*	7	.28
banish	10	.40	*brittle*	4	.16	*folly*	2	.08
pertinent	11	.44	*unequivocally*	2	.08	*rambunctious*	6	.24

Appendix 6. Main Study Results
D. GRG frequency of CQs over the three weeks of learning

1. CQ scores in week 1

Frequency of participants scores in graded reader CQs in the first week of learning

Target words	**Score**		
	zero	1	2
vicissitude	1	12	12
proprietorship	1	5	19
postulation	0	13	12
contentious	2	9	14
ingenious	1	10	14
new-fangled	0	3	22
perceptive	2	8	15
concur	1	7	17
elucidate	4	8	13
shrewd	2	7	16
trespass	0	12	13
contrive	0	5	20
detestable	2	7	16
banish	0	4	21
pertinent	0	12	13

2. CQ scores in week 2

Frequency of participants scores in graded reader CQs in the second week of learning

Target words	**Score**		
	zero	1	2
adroit	1	11	13
stupendous	1	6	18
misgiving	0	11	14
gargantuan	2	10	13
irksome	1	11	13
mystification	2	7	16

ebullient	3	8	14
idiosyncratic	2	8	15
acrid	4	8	13
obstinate	3	7	15
colossal	0	11	14
bonny	1	7	17
apprehensive	2	7	16
brittle	1	7	17
unequivocally	0	12	13

3. CQs scores in week 3

Frequency of participants scores in graded reader CQs in the third week of learning

Target words	**Score**		
	zero	1	2
grievous	0	15	10
cogent	0	5	20
toil	0	14	11
tedious	2	9	14
expeditious	2	10	13
astound	1	8	16
veritably	2	8	15
feeble	2	6	17
afflicted	5	9	11
sluggishly	3	11	11
petrify	0	15	10
lucid	1	9	15
perseverance	3	8	14
folly	2	9	14
rambunctious	1	14	10

Appendix 6. Main Study Results

E. Results and discussion of "*How does the retention of target words in lists A, B and C differ in the weekly immediate post-tests from the delayed post-test for the RMCG, the GRG, and the CG?"*

1- Results:

A comparison of LDT AS of the three groups in list A

This section will compare the AS for list A in the three groups The RMCG has a mean AS of 12.94 (SD = 1.324). Table 1 shows the means and standard deviations of AS of the RMCG in the four LDTs.

Table 1 *The AS of the RMCG of list A in the four LDTs*

LDT	**LDT 1**	**LDT 2**	**LDT 3**	**LDT 4**
list A LDT AS (max =15)	13.12	13.32	12.10	12.60
SD	1.01	.98	1.48	1.63

A Kruskal-Wallis H test is conducted to determine if there is a significant difference in LDT AS between the RMCG performance in the four LDTs. The Kruskal-Wallis H reveals that LDT AS values are not significant between the four LDTs for the RMCG, (χ^2 (3) = 3.454, p = .327).

The mean of the LDT AS of the GRG in list A is 11.24 (SD = 1.753). Table 2 shows the means and standard deviations of AS of the GRG in the four LDTs.

Table 2 *The AS of the GRG in list A in the four LDTs*

LDT	**LDT 1**	**LDT 2**	**LDT 3**	**LDT 4**
list A LDT AS (max = 15)	12.48	11.32	10.52	10.64
SD	1.19	1.74	2.00	1.287

A Kruskal-Wallis H test is conducted to determine if there are differences in LDT AS between the GRG performance in the four LDTs. The Kruskal-Wallis H test reveals that the difference between LDT AS for the GRG is significant, (χ^2 (3) = 20.248, p <.0001).

A Mann Whitney U test is conducted to compare the differences between the LDT AS of the GRG in list A in the weekly immediate post-tests and the delayed post-test. This analysis reveals that the difference between LDT 1 AS and LDT 2 AS of the GRG is significant where (U = 197.00, z = - 2.291, p = .022). Also, the difference between LDT 1 AS and LDT 4 AS of the GRG is significant (U = 92.50, z = - 4.355, p < .0001).

The CG mean of LDT AS is 5.97 (SD = 2.397). Table 3 shows the means and standard deviations of AS of the CG in the four LDTs.

Table 3 *The AS of the CG in list A in the four LDTs*

LDT	**LDT 1**	**LDT 2**	**LDT 3**	**LDT 4**
LDT AS	5.04	5.96	6.00	6.88

(max = 15)				
SD	2.70	2.47	1.77	2.31

A Kruskal-Wallis H test is conducted to determine if there are differences between the CG performance in four LDTs. The Kruskal-Wallis H test LDT AS reveals that the difference between the LDT AS of the CG in the four LDTs is not significant, (χ^2 (3) = 6.826, p = .078). Figure 1 shows the means of AS and standard deviation of the RMCG, GRG and CG and in list A in the four LDTs.

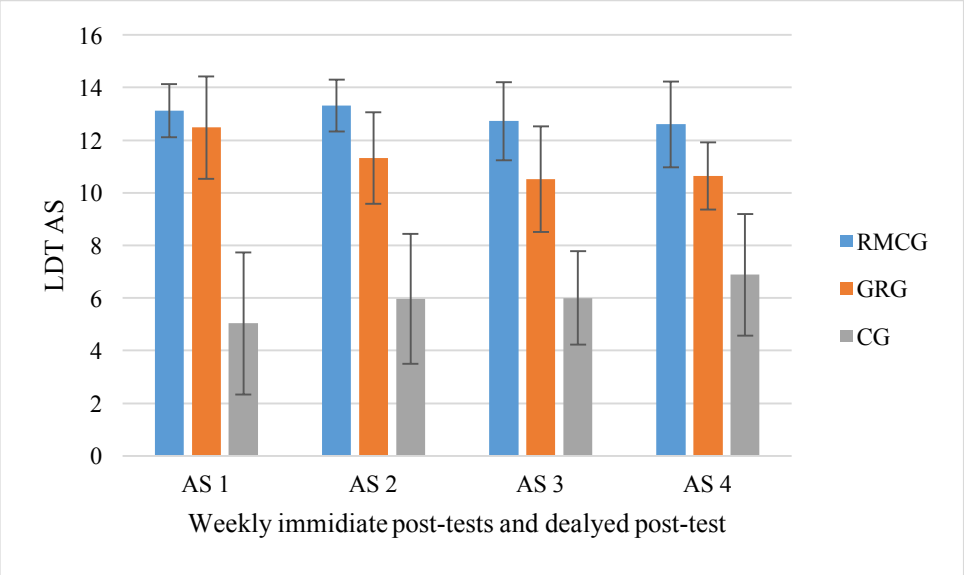


Figure 1 The LDT AS of RMCG, GRG and the CG in list A in the weekly immediate post-tests and the delayed post-test

A comparison of LDT AS of the three groups in list B

This section will compare the AS for list B in the three groups. The RMCG has a mean of 11.72 (SD= 1.69) in the LDT AS in the weekly immediate post-tests and the delayed post-test. Table 4 shows the means and standard deviations of AS of the RMCG in list B in the three LDTs.

Table 4 *The AS of the RMCG in list B in the three LDTs*

LDT	**LDT 2**	**LDT 3**	**LDT 4**
LDT AS (max = 15)	12.08	12.04	11.04
SD	1.60	1.61	1.71

A Kruskal-Wallis H test is conducted to determine if there are differences in AS between the RMCG performance in list B in the three LDTs. The Kruskal-Wallis H test reveals a significant difference between the RMCG performance in list B in the three LDTs, (χ^2 (2) = 7.50, p = .024).

A Mann Whitney U test is conducted to compare the differences between the LDT AS of the RMCG in list B in the weekly immediate post-tests and the delayed post-test. This analysis shows that the difference between LDT 3 AS and LDT 4 AS in list B is significant where (U = 203.50, z = -2.153, p = .030). The difference between LDT 2 AS and LDT 4 AS in list B is significant (U = 183.50, z = -2.540, p = .011).

The GRG has a mean of 10.12 points out of 15 (SD = 1.66) in LDT AS in the weekly immediate post-tests and the delayed post–test. Table 5 shows the means and standard deviations of AS of the GRG in the three LDTs.

Table 5 *The AS of the GRG in list B in the three LDTs*

LDT	**LDT 2**	**LDT 3**	**LDT 4**
LDT AS (max = 15)	10.32	9.72	10.32
SD	1.43	1.74	1.77

A Kruskal-Wallis H test is conducted to determine if there are differences in LDT AS between the GRG performance in list B in the three LDTs. The Kruskal-Wallis H test reveals no statistically significant difference between the three LDTs, (χ^2 (2) = 3.025, p = .220).

The CG has a mean of 6.23 points out of 15 in LDT AS (SD = 1.86) in the weekly immediate post-tests and the delayed post-test. Table 6 shows the means and standard deviations of AS of the CG in the three LDTs.

Table 6 *The AS of the CG in list B in the three LDTs*

LDT	**LDT 2**	**LDT 3**	**LDT 4**
LDT AS (max = 15)	6.31	5.88	6.52
SD	1.97	1.90	1.73

A Kruskal-Wallis H test is conducted to determine if there are differences in LDT AS between the CG performance in the three LDTs. The Kruskal Wallis-H test reveals no statistically significant difference between AS of the CG in the three LDTs, (χ^2 (2) = 1.722, p = .423). Figure 2 shows a comparison of LDT AS of the RMCG, the GRG and the CG in the weekly immediate post-tests and the delayed post-test.

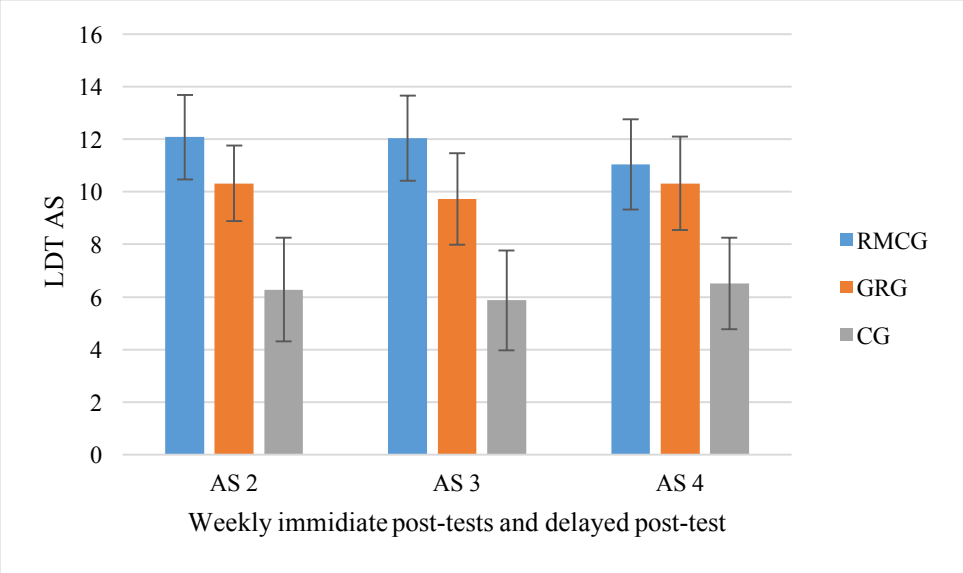


Figure 2 A comparison of the RMCG, GRG and CG LDT AS in list B in the weekly immediate post-tests and the delayed post-test.

A comparison of LDT AS of the three groups in list C

This section will compare the AS for list C in the three groups. The mean LDT AS of the RMCG in the last weekly immediate post-test and the delayed post-test is 10.58 points out of 15 (SD = 1.71). Table 7 shows the means and standard deviations of AS of the RMCG in the third and fourth LDT.

Table 7 *The AS of the RMCG in list C in LDT 3 and LDT 4*

LDT	**LDT 3**	**LDT 4**
LDT AS (max = 15)	10.76	10.40
SD	1.58	1.84

A Mann Whitney U test is conducted to compare the differences between the LDT AS of the RMCG in list C in LDT 3 and LDT 4. This analysis reveals that the difference or the RMCG between LDT 3 AS and LDT 4 AS is not significant, (U = 275.00, z = - .747, p = .455).

The mean LDT AS of the GRG in the last weekly immediate post-test and the delayed post-test is 10.42 points out of 15 (SD = 1.386). Table 8 shows the means and standard deviations of AS of the GRG in list C in LDT 3 and LDT 4.

Table 8 *The AS of the GRG in list C in LDT 3 and LDT 4*

LDT	**LDT 3**	**LDT 4**
LDT AS (max = 15)	10.16	10.68
SD	1.46	1.28

A Mann Whitney U test is conducted to compare the differences between the LDT AS of the GRG in LDT 3 and LDT 4. This analysis revealed that the difference between LDT 3 AS and LDT 4 AS of the GRG is not significant, (U = 257.50, z = -1.100, p = .271).

The mean LDT AS of the CG in the weekly immediate post-test and the delayed post-test is 6.24 points out of 15 (SD = 1.824). Table 9 shows means and standard deviations of AS of the CG in LDT 3 and LDT 4.

Table 9 *The AS of the CG in list C in LDT 3 and LDT 4*

LDT	**LDT 3**	**LDT 4**
LDT AS (max = 15)	5.52	6.96
SD	1.78	1.59

A Mann Whitney U test was conducted to compare the differences between the LDT AS of the CG in LDT 3 and LDT 4. This analysis revealed that the difference between LDT 3 AS and LDT 4 AS is significant, (U = 180.50, z = - 2.597, p = .009). Figure 3 shows a comparison between the three groups in LDT AS in LDT 3 and LDT 4.

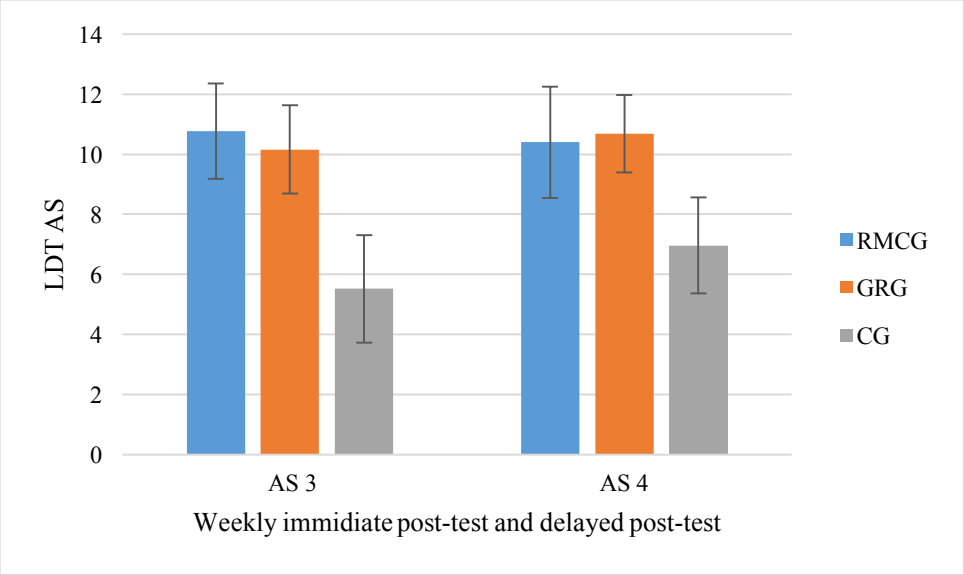


Figure 3 A comparison of LDT AS of the RMCG, GRG and CG in list C in weekly immediate post-test and the delayed post-test.

A comparison of RT of the three groups in list A

This section will compare the RTs for list Ain the three groups. The mean RTs of the RMCG is 2093.70 *ms* (SD = 364.184) in the weekly immediate post- tests and the delayed post-test. The RTs means and standard deviations of the RMCG in the four LDTs are presented in Table 10.

Table 10 *The RT of the RMCG for list A in the four LDTs*

LDT	**LDT 1**	**LDT 2**	**LDT 3**	**LDT 4**
list A LDT RT (*ms*)	2354.97	2224.75	1710.87	2084.22
SD	256.81	388.67	260.25	276.14

A Kruskal-Wallis H test is conducted to determine if there are differences in RTs between the RMCG performance in the four LDTs. The Kruskal-Wallis H test reveals a statistically significant difference of the RT for the RMCG in list A in the four LDTs, (χ^2 (3) = 43.426, $p < .0001$). A Mann Whitney U test is conducted to compare the differences between the LDT RT of the RMCG in the weekly immediate post-tests and the delayed post-test. This analysis reveals a significant difference between RT 2 and RT 3 where ($U = 63.00$, $z = -4.841$, $p < .0001$). In addition, it reveals a significant difference between RT3 and RT 4 where ($U = 104.00$, $z = -4.045$, $p < .0001$) and reveals a significant difference between RT 1 and RT 4 where ($U = 142.00$, $z = -3.308$, $p = .001$).

The mean of the RT of the GRG is 2210.80 *ms* (SD = 494.872) in the weekly immediate –post-tests and the delayed post-test. The RTs means and standard deviations of the GRG in the four LDTs is presented in Table 11.

Table 11 *The RT of the GRG in list A in the four LDTs*

LDT	**LDT 1**	**LDT 2**	**LDT 3**	**LDT 4**
LDT RT (*ms*)	2756.32	2447.11	1926.34	1713.43
SD	265.26	241.59	271.34	312.57

A Kruskal-Wallis H test is conducted to determine if there are difference in RTs in the four LDTs of the GRG. The Kruskal-Wallis H test reveals that there is a significant difference between the RTs in the four LDTs, (χ^2 (3) = 72.479, p < .0001). A Mann Whitney U test is conducted to compare the differences between the RT of the GRG in the weekly immediate post-tests and the delayed post-test. This analysis shows that the difference between RT 1 and RT 2 is significant, (U = 128.00, z = -453.00, p < .0001). In addition, the difference between RT 2 and RT 3 is significant, (U = 45.00, z = -370.00, p < .0001) and the difference between RT 3 and RT 4 is significant, (U = 203.00, z = -2.125, p= .034). The difference between RT 1 and RT 4 is significant, (U = .000, z = -6.063, p < .0001).

The RTs of the CG in the weekly immediate post-tests and the post-test is 2092.30 *ms* (SD = 808.96). The RTs means and standard deviations of the CG in the four LDTs are presented in Table 12.

Table 12 *The RT of the CG in list A in the four LDTs*

LDT	**LDT 1**	**LDT 2**	**LDT 3**	**LDT 4**
LDT RT (*ms*)	2824.10	2286.00	1749.88	1509.22
SD	949.72	616.12	489.72	323.80

A Kruskal-Wallis H test is conducted to determine if there are differences in RTs between the CG performance the four LDTs. The Kruskal-Wallis H test reveals a statistically significant difference between the RTs in the four LDTs, (χ^2 (3) = 49.89, p < .0001). A Mann Whitney U test is conducted to compare the differences between the RTs in the weekly immediate post-tests and the delayed post-test. This analysis reveals a statistically significant difference between RT 1 and RT 2 where (U = 199.500, z = -2.193, p = .028). In addition, it reveals a statistically significant difference between RT 2 and RT 3 where (U = 109.50, z = -3.939, p < .0001) and statistically significant difference between RT 3 and RT 4 where (U = 176.00, z = -2.648, p = .008). The difference between RT 1 and RT 4 is significant where (U = 40.00, z = -5.287, p < .0001). Figure 4 shows a comparison of RTs in list A for the three groups.

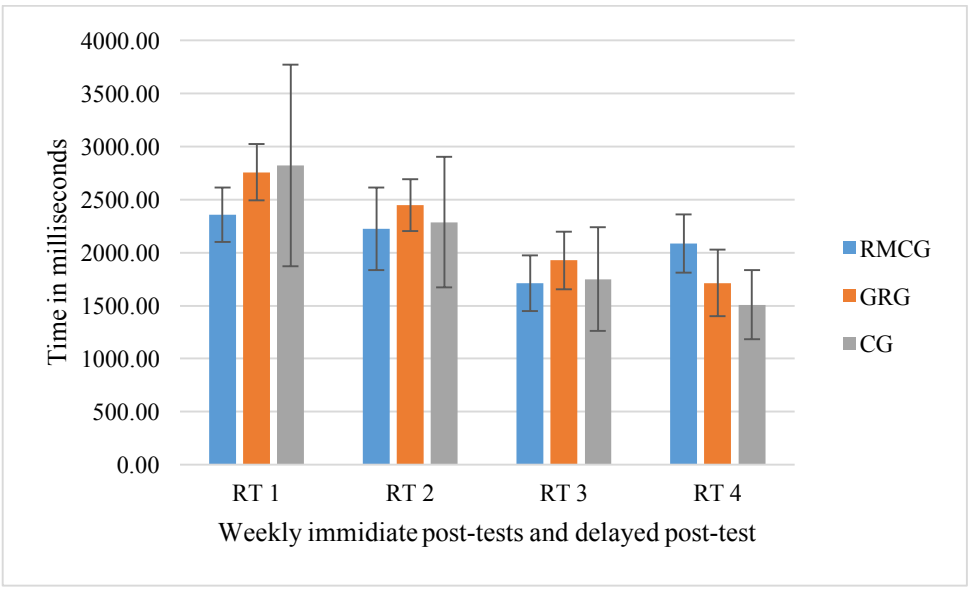


Figure 4 The LDT RT in list A in the weekly immediate post-tests and the delayed post-test

A comparison of RT of the three groups in list B

This section will compare the RTs for list B in the three groups. The mean RT of the RMCG in the weekly immediate post-tests and the delayed post-test is 2095.627 *ms* (SD = 284.77). The RTs means and standard deviations of the RMCG in the three LDTs are presented in Table 13.

Table 13 *The RT of the RMCG in list B in the three LDTs*

LDT	**LDT 2**	**LDT 3**	**LDT 4**
LDT RT (*ms*)	2250.30	2116.88	1919.69
SD	291.01	258.89	201.73

A Kruskal-Wallis H test is conducted to determine if there are differences in RT between the RMCG performance in list B in the three LDTs. The Kruskal-Wallis H test reveals a significant difference in the RT of the RMCG in the three LDTs, (χ^2 (2) = 18.075, $p < .0001$). A Mann Whitney U test is conducted to compare the differences between the RT of the RMCG in the weekly immediate post-tests and the delayed post-test. This analysis reveals that the difference between RT 2 and RT 3 is significant, ($U = 202.00$, $z = -2.144$, $p = .032$). In addition, the difference between RT 3 and RT 4 is significant, ($U = 171.00$, $z = -2.746$, $p = .006$) and the difference between RT 2 and RT 4 is significant, ($U = 112.00$, $z = -3.890$, $p < .0001$).

The mean RT of the GRG in the weekly immediate post-tests and the delayed post-test is 2235.86 *ms* (SD = 426.28). The RTs means and standard deviations of the GRG in the three LDTs are presented in Table 14.

Table 14 *The RT of the GRG in list B in the three LDTs*

LDT	**LDT 2**	**LDT 3**	**LDT 4**
LDT RT (*ms*)	2378.86	2509.64	1819.08
SD	343.21	368.34	162.75

A Kruskal-Wallis H test was conducted to determine if there are differences in RTs between the GRG performance in list the three LDTs. The Kruskal-Wallis H test reveals that RT values are significantly different between list B in the three LDTs, (χ^2 (2) = 40.357, $p < .0001$). A Mann Whitney U test is conducted to compare the differences between the RTs of the GRG in the weekly immediate post-tests and the delayed post-test. This analysis reveals that the difference between RT 3 and RT 4 is significant, ($U = 39.00$, $z = -5.307$, $p < .0001$). In addition, the difference between RT 2 and RT 4 is significant ($U = 27.00$, $z = -5.540$, $p < .0001$).

The mean RT of the CG in the weekly immediate post-tests and the delayed post-test is 1932.56 *ms* (SD = 327.51). The RTs means and standard deviations of the CG in list B in the three LDTs are presented in Table 15.

Table 15 *The RT of the CG in list B in the three LDTs*

LDT	**LDT 2**	**LDT 3**	**LDT 4**
LDT RT (*ms*)	2126.69	1753.83	1917.08
SD	294.87	327.89	251.96

A Kruskal-Wallis H test is conducted to determine if there are differences in RTs between the CG performance in the three LDTs. The Kruskal-Wallis H test reveals a statistically significant difference between list B in the three LDTs, (χ^2 (2) = 17.856, $p < .0001$). A Mann Whitney U test is conducted to compare the differences between the RTs of the CG in the weekly immediate post-tests and the delayed post-test. This analysis reveals that the difference between RT 2 and RT 3 is significant where ($U = 113.00$, $z = -3.871$, $p < .0001$). In addition, the difference between RT 3 and RT 4 is significant where ($U = 203.00$, $z = -2.125$, $p = .034$), and the difference between RT 2 and RT 4 is significant where ($U = 127.00$, $z = -2.726$, $p = .006$). Figure 5 shows a comparison of the three groups RTs means and standard deviations in list B in the weekly immediate post-tests and the delayed post-test.

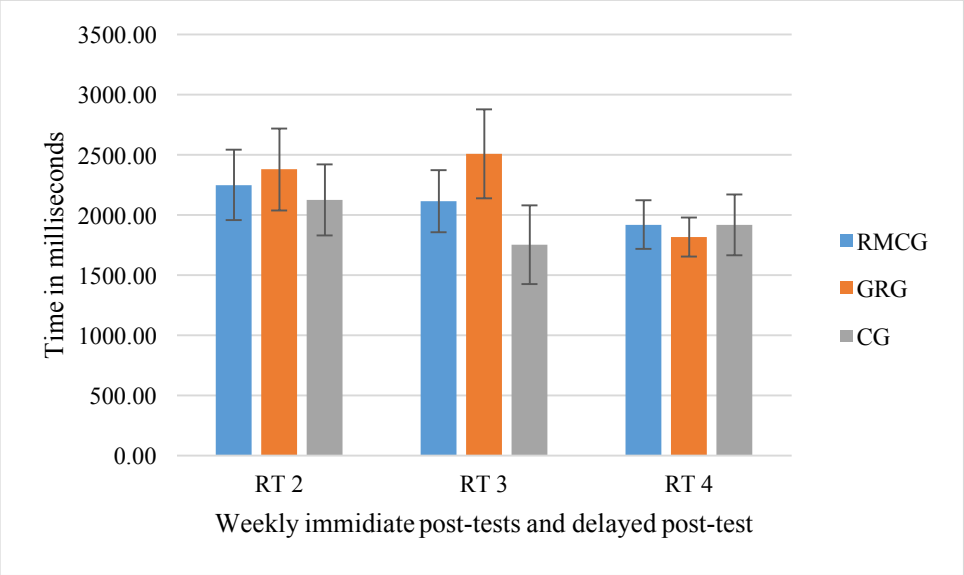


Figure 5 The LDT RT in list B in the weekly immediate post-tests and the delayed post-test

A comparison of the RTs in list C for all groups

This section will compare the RTs for list C in the three groups. The mean RT of the RMCG in the last weekly immediate post-test and the delayed post–test is 2413.00 *ms* (SD = 590.816). The RTs means and standard deviations of the RMCG in the LDT 3 and LDT 4 are presented in Table 16. A Mann Whitney U test reveals that the difference between RT 3 and RT 4 is not significant where (U = 228.00, z = -1.640, p = .101).

Table 16 *The RT of the RMCG in list C in the LDT 3 and LDT 4*

LDT	**LDT 3**	**LDT 4**
LDT RT (*ms*)	2611.00	2216.05
SD	755.89	245.05

The mean RTs of the GRG in LDT 3 and LDT 4 is 2121.18 *ms* (SD = 472.69). The RTs means and standard deviations of the GRG in the LDT 3 and LDT 4 are presented in Table 17.

Table 17 *The RT of the GRG for list C in the LDT 3 and LDT 4*

LDT	**LDT 3**	**LDT 4**
LDT RT (*ms*)	2423.38	1818.99
SD	453.45	245.60

A Mann Whitney U test is conducted to compare the differences between the LDT RT of the GRG in the last weekly immediate post-test and the delayed post-test. This analysis reveals that the difference between RT 3 and RT 4 is significant, (U = 74.00, z = - 4.628, p < .0001).

The mean of RT of the CG in the last weekly immediate post-test and the delayed post-test is 1764.36 *ms* (SD = 388.48). The RTs means and standard deviations of CG in the LDT 3 and LDT 4 are presented in Table 18.

Table 18 *The RT of the CG for list C in the LDT 3 and LDT 4*

LDT	**LDT 3**	**LDT 4**
LDT RT (*ms*)	1891.98	1636.74
SD	421.78	310.32

A Mann Whitney U test reveals that the difference between LDT 2 RT and LDT 3 RT in list C for the CG is significant, (U = 208.00, z = - 2.028, p = .043). Figure 6 shows the RT of the three groups in list C in LDT 3 and LDT 4.

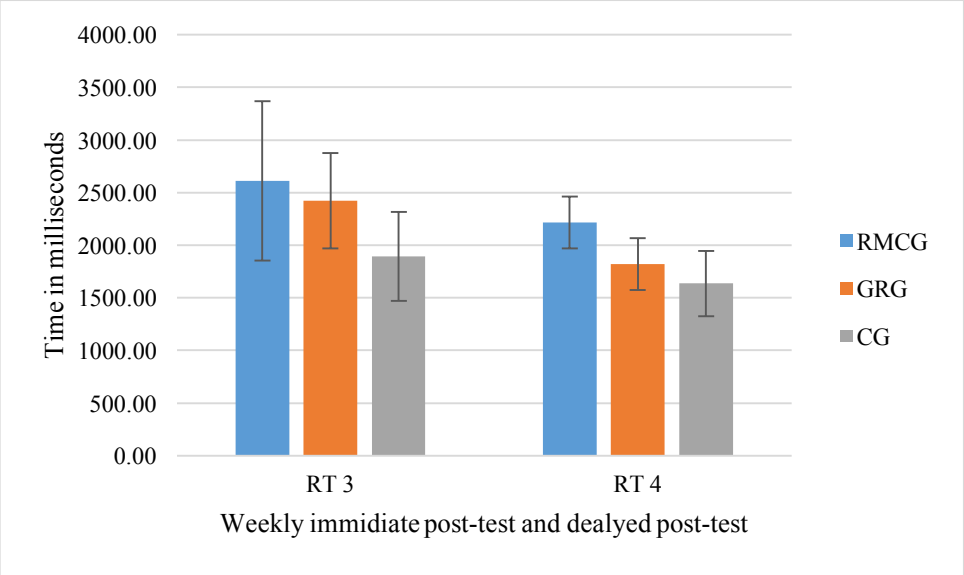


Figure 6 The LDT RT in list C in the weekly immediate post-tests and the delayed post-test

A comparison of the GFT scores of the three groups in list A

This section will compare the GFT scores for list A in the three groups. The mean of GFT scores of the RMCG in the weekly immediate post-tests and the delayed post-test is 12.20 out of 15 (SD = 1.979). The GFT scores of the RMCG is in Table 19.

Table 19 *The GFT scores means and standard deviations of the RMCG in list A*

GFT	**GFT 1**	**GFT 2**	**GFT 3**	**GFT 4**
GFT score (max = 15)	11.68	12.72	10.96	9.52
SD	2.11	1.72	1.90	2.12

A Kruskal-Wallis H test is conducted to determine if there are differences in GFT scores between the RMCG performance in the four GFTs. The Kruskal-Wallis H test reveals a statistically significant difference in the four GFTs, (χ^2 (3) = 26.623, p < .0001). A Mann Whitney U test is conducted to compare the differences between GFT scores of the RMCG in the weekly immediate post-tests and the delayed post-test. This analysis reveals that the difference between GFT 2 scores and GFT 3 scores is significant where (U = 148.50, z = - 3, p = .001). In addition, the difference between GFT 3 scores and GFT 4 scores is significant where (U = 186.00, z = -2.483, p = .013) and the difference between GFT 1 scores and GFT 4 scores is significant (U = 141.50, z = -3.345, p = .001).

The mean GFT scores of the GRG for in the weekly immediate post-tests and the delayed post-test is 7.92 (SD = 1.860). The GFT scores means and standard deviation is presented in Table 20.

Table 20 *The GFT scores means and standard deviations of the GRG for list A*

GFT	**GFT 1**	**GFT 2**	**GFT 3**	**GFT 4**
GFT score Max = 15	8.92	6.92	6.36	5.44
SD	1.57	1.57	1.68	1.41

A Kruskal-Wallis H test is conducted to determine if there are differences in GFT scores between the GRG performance in the four GFTs. The Kruskal-Wallis H test reveals significant difference between GFT scores in the four GFTs, (χ^2 (3) = 41.757, $p < .0001$). A Mann Whitney U test is conducted to compare the differences between the GFT scores of the GRG in the weekly immediate post-tests and the delayed post-test. This analysis reveals that the difference between GFT 1 scores and GFT 2 scores is significant, ($U = 119.00$, $z = -3.813$, $p < .0001$). In addition, the difference between GFT 3 scores and GFT 4 scores is significant where ($U = 199.50$, $z = -2.248$, $p = .025$) and the difference between GFT 1 scores and GFT 4 scores is significant where ($U = 29.00$, $z = -5.546$, $p < .0001$). Figure 5.21 shows the means and standard deviations of the GFT scores for the RMCG and the GRG for list A in the weekly immediate post-tests and the delayed post-test.

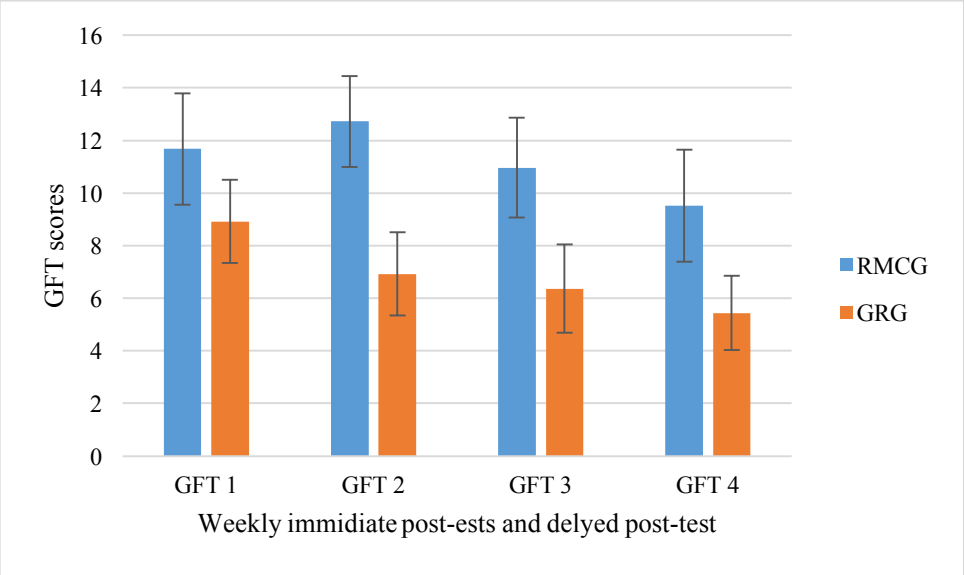


Figure 7 A comparison of the GFT scores of the RMCG and the GRG in list A in the weekly immediate post-tests and the delayed post-test

A comparison of the GFT scores of the three groups in list B

This section will compare the GFT scores for list B in the three groups. The mean GFT score for the RMCG in the weekly immediate post-tests and the delayed post-test is 9.88 (SD = 2.01). Table 21 shows the means and tendered deviations for the GFT scores of the RMCG in list B.

Table 21 *The GFT scores means and standard deviations of the RMCG in list B*

GFT	**GFT 2**	**GFT 3**	**GFT 4**
GFT score (max = 15)	9.64	11.08	8.92

SD	1.75	1.77	1.95

A Kruskal-Wallis H test is conducted to determine if there are differences in GFT scores between the RMCG performance in the three GFTs. The Kruskal Wallis H test reveals that GFT scores of the RMCG are significantly different in the three GFTs, (χ^2 (2) = 15.902, p < .0001). A Mann Whitney U test is conducted to compare the differences between the GFT scores of the RMCG in the weekly immediate post-tests and the delayed post-test. This analysis reveala that the difference between GFT 2 scores and GFT 3 scores is significant where (U = 162.50, z = -2.95, p = .004). In addition, the difference between GFT 3 scores and GFT 4 scores is significant where (U = 129.50, z = - 3.595, p < .0001).

The mean GFT scores of the GRG in the weekly immediate post-tests and the delayed post-test is 6.9 (SD = 3.23). Table 22 shows the GFT scores means and standard deviations of the GRG.

Table 22 *The GFT scores means and standard deviations of the GRG for list B*

GFT	**GFT 2**	**GFT 3**	**GFT 4**
GFT score (max = 15)	10.24	5.60	4.88
SD	2.80	2.00	1.66

A Kruskal-Wallis H test is conducted to determine if there are differences in GFT scores between the performance of the GRG in the three GFTs. The Kruskal-Wallis H test reveals that GFT scores for list B is significantly different between the three GFTs, (χ^2 (2) = 38.894, p < .0001). A Mann Whitney U test is conducted to compare the differences between the GFT scores of the GRG in the weekly immediate post-tests and the delyed post-test. This analysis reveals a significant difference between the scores of GFT 2 and GFT 3 where (U = 52.00, z = - 5.084, p < .0005). In addition, the it reveals a significant difference between the scores of GFT 1 and GFT 4 where (U = 29.00, z = -5.525, p < .0001). Figure 8 shows a comparison of the GFT scores of the RMCG and the GRG in the weekly immediate post-tests and the delayed post-test.

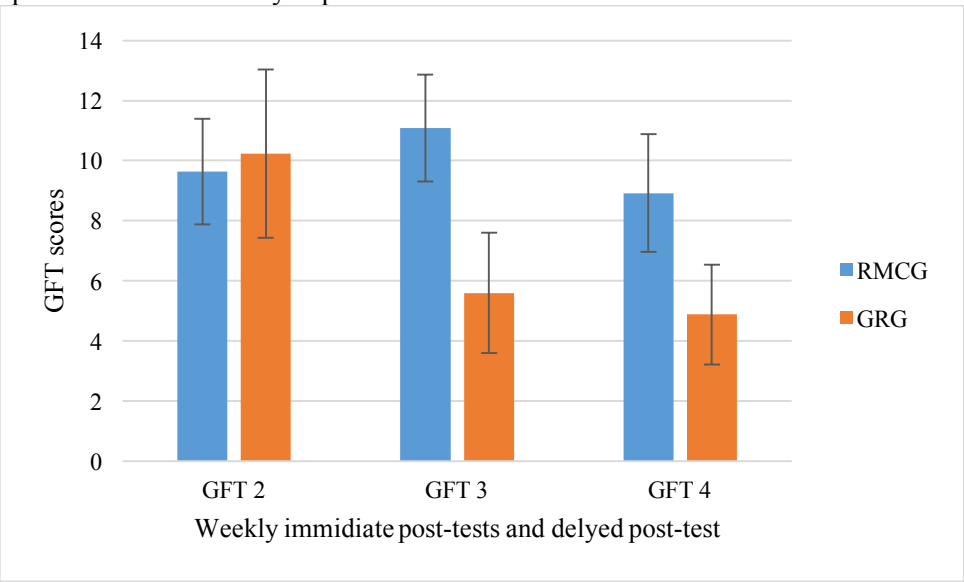


Figure 8 A comparison of GFT scores of the RMCG and the GRG in list B in the weekly immediate post-tests and the delayed post-test

A comparison of the GFT scores of the three groups in list C

This section will compare the GFT scores for list C in the three groups. The mean GFT score of the RMCG in the last weekly immediate post-test and the delayed post-test is 7.72 (SD = 2.13). Table 23 shows the GFT scores means and standard deviations of the RMCG.

Table 23 *The GFT scores means and standard deviations of the RMCG for list C*

GFT	**GFT 3**	**GFT 4**
GFT score (max = 15)	8.44	7.00
SD	2.10	1.95

A Mann Whitney U test is conducted to compare the differences between the GFT scores of the RMCG in the last weekly immediate post-tests and the delayed post-test. This analysis reveals that the difference between GFT 3 scores and GFT 4 scores is significant where (U = 198.50, z = - 2.233, p = .026).

The mean GFT scores of the GRG for list C in the last weekly immediate post-tests and the delayed post-test is 5.80 (SD = 2.03). Table 24 shows the GFT scores means and standard deviations of the GRG in GFT 3 and GFT 4.

Table 24 *The GFT scores means and standard deviations of the GRG for list C*

GFT	**GFT 3**	**GFT 4**
GFT score Max = 15	6.76	4.84
SD	1.96	1.62

A Mann Whitney U test is conducted to compare the differences between the GFT scores of the GRG in the last weekly immediate post-test and the delayed post-test. This analysis reveals that the difference between GFT 3 scores and GFT 4 scores in list C for the GRG is significant where (U = 137.00, z = - 3.455, p = .001). Figure 9 shows a comparison of the GFT scores of the RMCG and the GRG in the GFT 3 and GFT 4.

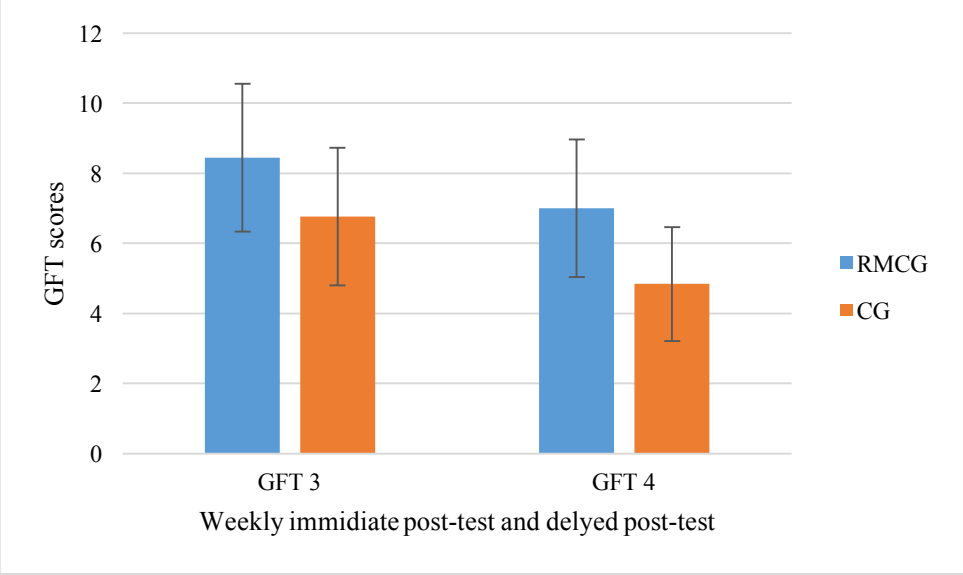


Figure 9 A comparison of the GFT scores of the RMCG and the GRG in list C in the last weekly immediate post-test and the delayed post-test

To investigate the target words retention independently, the percentage of participant recognition in LDT and production in GFT for each word in three lists (i.e. A, B and C) are calculated. Also, the mean of RTs for each word are calculated to compare the speed of response of each word in the three groups.

An analysis of the three groups recognition of each target word in list in list A

The percentage of the participant word recognition for list A in the weekly immediate post-tests and the delayed post-test is calculated for the three groups. This analysis will show the improved performance of the three groups in list A from LDT 1 to LDT 4. It will also highlight the fluctuated number of participant recognition of some of the target word. then it will show what target words have lower number of participant recognition in LDT 4 after two weeks from learning and what target words have higher number of participant recognition in LDT 4 after two weeks from learning. Following this is a highlight of target word with the highest number of participant recognition and of target word with the lowest number of participant recognition.

The RMCG recognition of target word in list A

Target word that has a decreasing number of participants' recognition from LDT 1 to LDT 4 for the RMCG are *ingenious, new-fangled, perceptive* and *concur.* Target word that has a fluctuated number of participant recognition from LDT 1 to LDT 4 are *vicissitude proprietorship, postulation, contentious, elucidate, shrewd, trespass, contrive, detestable, banish,* and *pertinent.*

Target word that has a lower number of participants' recognition after two weeks from learning in the delayed post-test than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *vicissitude, proprietorship, contentious, ingenious, new-fangled, perceptive, concur* and *detestable*. Target word that has a higher number of participant recognition after two weeks from learning than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *elucidate, shrewd, trespass, contrive* and *pertinent*.

All of the participants (i.e. 100% of the participants) in the RMCG recognise the word *new-fangled* in LDT 1 and the word *trespass* in LDT 2. The lowest recognition number is for the words *postulation, detestable* and *pertinent* in LDT 1 with 76% of the participants' recognition. Finally, the number of participant recognition is higher in LDT 2 than other weekly immediate post-tests or delayed post-test.

The GRG recognition of target word in list A

Target word that has a decreasing number of participants' recognition from LDT 1 to LDT 4 for the GRG are *concur* and *contrive.* The rest of the target word in list A has a fluctuated number of participant recognition from LDT 1 to LDT 4.

Target word that has a lower number of participants' recognition after two weeks from learning in the delayed post-test than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *vicissitude, contentious, perceptive, contrive* and *pertinent.* Target word that has a higher number of participant recognition after two weeks from learning than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *postulation, ingenious, new-fangled, elucidate, trespass detestable* and *banish.*

The highest number of participant recognition (i.e. 96 % of the participants) in the GRG is for the target word *trespass* in LDT 2. The lowest recognition number (i.e. 56%) is for the words *banish*

and *pertinent* in LDT 3 and LDT 4 respectively. Finally, the number of participant recognition is higher in LDT 1 than other weekly immediate post-tests or delayed post-test.

The CG recognition of target word in list A

Unlike the RMCG and the GRG, the CG has no target word with decreasing number of participant recognition from LDT 1 to LDT 4. Conversely, the target word *vicissitude, new-fangled, concur, shrewd, banish* and *pertinent* have an increasing number of participant recognition from LDT 1 to LDT 4. The rest of target word in list A have a fluctuated number of participant recognition from LDT 1 to LDT 4.

Target word that has a lower number of participants' recognition after two weeks without testing to the delayed post-test are *perceptive, elucidate* and *detestable.* Target word that has a higher number of participant recognition after two weeks without testing in the delayed post-test are the rest of the target word in list A.

The highest number of participant recognition (i.e. 64 % of the participants) in the CG is for the target word *Banish* in LDT 4. The lowest recognition number (i.e. 20%) is for the word *contentious* in LDT 1. Finally, the number of participant recognition is higher in LDT 4 than other weekly immediate post-tests.

An analysis of the three groups recognition of each target word in list B

The percentage of the participant word recognition for list B in the weekly immediate post-tests and the delayed post-test is calculated for the three groups. This analysis will show the improved performance of the three groups in list B from LDT 2 to LDT 4. It will also highlight the fluctuated number of participant recognition of some of the target word. then it will show what target word have lower number of participant recognition in LDT 4 after two weeks from learning and what target words have higher number of participant recognition in LDT 4 after two weeks from learning. Following this is a highlight of the target word with the highest number of participant recognition and of target word with the lowest number of participant recognition.

The RMCG recognition of target word in list B

Target word that has a decreasing number of participants' recognition from LDT 2 to LDT 4 for the RMCG are *gargantuan, irksome, ebullient, acrid, obstinate, apprehensive, brittle* and *unequivocally.* Target word that has a fluctuated number of participant recognition from LDT 2 to LDT 4 are *adroit, stupendous, misgiving, mystification, colossal* and *bonny.*

Target word that has a lower number of participants' recognition after two weeks from learning in the delayed post-test than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *adroit, stupendous, misgiving, gargantuan, irksome, mystification, ebullient, acrid, obstinate, apprehensive, brittle* and *unequivocally.* Target word that has a higher number of participant recognition after two weeks from learning than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *idiosyncratic, colossal* and *bonny.*

The highest number of the participant recognition is 92% for the target word *idiosyncratic* and *colossal* in LDT 1 and the word *bonny* in LDT 4. The lowest recognition number is for the words *apprehensive* in LDT 4 with 60% of the participant recognition. Finally, the number of participant recognition is higher in LDT 2 than other weekly immediate post-tests or the delayed post-test.

The GRG recognition of target word in list B

Target word that has a decreasing number of participants' recognition from LDT 2 to LDT 4 for the GRG are *adroit, misgiving ebullient, idiosyncratic, acrid, obstinate, bonny, apprehensive,* brittle

and unequivocally. The rest of the target word in list B has a fluctuated number of participant recognition from LDT 2 to LDT 4.

Target word that has a lower number of participants' recognition after two weeks from learning in the delayed post-test than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *adroit, misgiving, mystification, ebullient idiosyncratic, acrid, obstinate, bonny* and *brittle.* Target word that has a higher number of participant recognition after two weeks from learning than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *stupendous, gargantuan, irksome, gargantuan, irksome* and *colossal.*

The highest number of participant recognition (i.e. 88 % of the participants) in the GRG is for the target word *brittle* in LDT 2. The lowest recognition number (i.e. 44%) is for the word *stupendous* in LDT 3. Finally, the number of participant recognition is higher in LDT 2 than other weekly immediate post-tests or delayed post-test.

The CG recognition of target word in list B

Target word with decreasing number of participant recognition from LDT 2 to LDT 4 are *gargantuan, obstinate colossal* and *unequivocally*. Conversely, the target word *irksome*, *idiosyncratic* and *acrid* have an increasing number of participant recognition from LDT 2 to LDT 4. The rest of target word in list B have a fluctuated number of participant recognition from LDT 2 to LDT 4.

Target word that has a lower number of participants' recognition after two weeks without testing to the delayed post-test are *ebullient, colossal, apprehensive* and *brittle.* Target word that has a higher number of participant recognition after two weeks without testing in the delayed post-test are the rest of the target word in list B.

The highest number of participant recognition (i.e. 64 % of the participants) in the CG is for the target word *stupendous* in LDT 4. The lowest recognition number (i.e. 28%) is for the word *ebullient* in LDT 4. Finally, the number of participant recognition is higher in LDT 4 than other weekly immediate post-tests.

An analysis of the three groups recognition of each target word in list C

The percentage of the participant word recognition for list C in the last weekly immediate post-test and the delayed post-test is calculated for the three groups. This analysis will show the improved performance of the three groups in list C from LDT 3 to LDT 4. This analysis will show what target word have lower number of participant recognition in LDT 4 after two weeks from learning and what target words have higher number of participant recognition in LDT 4 after two weeks from learning. Following this is a highlight of the target word with the highest number of participant recognition and of target word with the lowest number of participant recognition.

The RMCG recognition of target word in list C

Target word that has a lower number of participants' recognition after two weeks from learning in the delayed post-test than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *grievous, cogent, toil, astound, veritably, feeble, lucid, perseverance, folly* and *rambunctious.* Target word that has a higher number of participant recognition after two weeks from learning than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are the rest of the target word except for the target word *afflicted* that does not change in LDT 4 from LDT 3.

The highest number of participant recognition (i.e. 88 % of the participants) in the RMCG is for the target word *folly* and *rambunctious* in LDT 3. The lowest recognition number (i.e. 56%) is for the word *petrify* in LDT 3. Finally, the number of participant recognition is higher in LDT 3 than the delayed post-test.

The GRG recognition of target word in list C

Target word that have a lower number of participants' recognition after two weeks from learning in the delayed post-test than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) are *grievous, astound, veritably, afflicted, sluggishly* and *rambunctious.* Target word that have a higher number of participant recognition after two weeks from learning than the number of participant recognition in the last weekly immediate post-test (i.e. LDT 3) *cogent, toil, tedious, expeditious, petrify, lucid, perseverance* and *folly.*

The highest number of participant recognition (i.e. 88 % of the participants) in the GRG is for the target word *sluggishly* in LDT 3. The lowest recognition number (i.e. 52%) is for the word *petrify* in LDT 3. Finally, the number of participant recognition is higher in LDT 4 than the last weekly immediate post-tests.

The CG recognition of target word in list C

Target word that has a lower number of participants' recognition after two weeks without testing to the delayed post-test are *afflicted* and *rambunctious.* Target word that has a higher number of participant recognition after two weeks without testing in the delayed post-test are the rest of the target word in list C.

The highest number of participant recognition (i.e. 64 % of the participants) in the CG is for the target word *veritably* in LDT 4. The lowest recognition number (i.e. 24%) is for the word *expeditious* in LDT 3. Finally, the number of participant recognition is higher in LDT 4 than the last weekly immediate post-tests.

An analysis of the RMCG and the GRG production of each target word in list A

The percentage of the participant word production for list A in the weekly immediate post-tests and the delayed post-test is calculated for the RMCG and the GRG. This analysis will show the improved performance of the three groups in list A from GFT 1 to GFT 4. It will also highlight the fluctuated number of participant production of some of the target word. Then it will show what target word have lower number of participant production in GFT 4 after two weeks from learning and what target word have higher number of participant production in GFT 4 after two weeks from learning. Following this is a highlight of target word with the highest number of participant production and target word with the lowest number of participant production.

The RMCG production of target word in list A

Target word that has a decreasing number of participant production from GFT 1 to GFT4 for the RMCG are *vicissitude, postulation, ingenious, shrewd* and *banish.* The rest of the target word in list A have a fluctuated number of participant production from GFT 1 to GFT 4.

Target word that has a higher number of participant production after two weeks from learning in the delayed post-test than the number of participant production in the last weekly immediate post-test (i.e. GFT 3) is the word *new-fangled.* The rest of the target word in list A have lower number of participant production after two weeks from learning than the number of participant production in the last weekly immediate post-test (i.e. GFT 3).

The highest number of participant production (i.e. 100 % of the participants) in the RMCG is for the target word *banish* in GFT 1. The lowest participant production number (i.e. 40%) is for the

word *postulation* in GFT 4. Finally, the mean number of participant production is higher in GFT 2 than other weekly immediate post-tests or delayed post-test.

The GRG production of target word in list A

Target word that has a decreasing number of participant production from GFT 1 to GFT4 for the GRG are *vicissitude, proprietorship, postulation, contentious, ingenious, perceptive, elucidate, shrewd, detestable, banish* and *pertinent.* The rest of the target word in list A have a fluctuated number of participant production from GFT 1 to GFT 4.

Target word that has a higher number of participant production after two weeks from learning in the delayed post-test than the number of participant production in the last weekly immediate post-test (i.e. GFT 3) is the word *new-fangled* and *concur.* The rest of the target word in list A have lower number of participant production after two weeks from learning than the number of participant production in the last weekly immediate post-test (i.e. GFT 3).

The highest number of participant production (i.e. 88 % of the participants) in the GRG is for the target word *banish* in GFT 1. The lowest participant production number (i.e. 12%) is for the word *postulation* in GFT 4. Finally, the mean number of participant production is higher in GFT 1 than other weekly immediate post-tests or delayed post-test.

An analysis of the RMCG and the GRG production of list B

The percentage of the participant word production for list B in the weekly immediate post-tests and the delayed post-test is calculated for the RMCG and the GRG. This analysis will show the improved performance of the three groups in list B from GFT 2 to GFT 4. It will also highlight the fluctuated number of participants' production of some of the target word. Then it will show what target word have lower number of participant production in GFT 4 after two weeks from learning and what target word have higher number of participant production in GFT 4 after two weeks from learning. Following this is a highlight of target word with the highest number of participant production and target word with the lowest number of participant production.

The RMCG production of target word in list B

Target word that has a decreasing number of participant production from GFT 2 to GFT4 for the GRG are *misgiving, apprehensive* and *brittle.* The rest of the target word in list A have a fluctuated number of participant production from GFT 2 to GFT 4.

All target word in list A have lower number of participant production after two weeks from learning in comparison to the number of participant production in the last weekly immediate post-test (i.e. GFT 3).

The highest number of participant production (i.e. 88 % of the participants) in the RMCG is for the target word *bonny* in GFT 3. The lowest participant production number (i.e. 40%) is for the word *brittle* in GFT 4. Finally, the mean number of participant production is higher in GFT 3 than other weekly immediate post-tests or delayed post-test.

The GRG production of target word in list B

All target word in list B except the word *gargantuan* have a decreasing number of participant production from GFT 1 to GFT 4. All target word except the word *gargantuan* has a lower number of participant production after two weeks from learning in the delayed post-test than the number of participant production in the last weekly immediate post-test (i.e. GFT 3.

The highest number of participant production (i.e. 88 % of the participants) in the GRG is for the target word *irksome* in GFT 2. The lowest participant production number (i.e. 12%) is for the word

mystification in GFT 3 and GFT 4. Finally, the mean number of participant production is higher in GFT 2 than other weekly immediate post-tests or delayed post-test.

of the RMCG and the GRG production of each target word in list C

The RMCG production of target word in list C

All target word except the words *tedious, expeditious* and *perseverance* have a lower number of participant production after two weeks from learning in the delayed post-test than the number of participant production in the last weekly immediate post-test (i.e. GFT 3). The target word *tedious, expeditious* and *perseverance* have similar GFT scores in GFT 3 and GFT 4. The highest number of participant production (i.e. 76 % of the participants) in the RMCG is for the target word *astound* in GFT 3. The lowest participant production number (i.e. 28%) is for the word *afflicted* in GFT 4. Finally, the mean number of participant production is higher in GFT 3 than the delayed post-test.

The GRG production of target word in list C

All target word except the words *tedious* and *expeditious* have a lower number of participant production after two weeks from learning in the delayed post-test than the number of participant production in the last weekly immediate post-test (i.e. GFT 3). The target word *tedious* and *expeditious* have similar GFT scores in GFT 3 and GFT 4. The highest number of participant production (i.e. 64 % of the participants) in the GRG is for the target word *tedious* and *sluggishly* in GFT 3. The lowest participant production number (i.e. 28%) is for the word *afflicted* in GFT 4. Finally, the mean number of participant production is higher in GFT 3 than the delayed post-test.

2- Discussion:

The data clearly shows that, as expected, the RMCG had no difficulty in acquiring almost all the 15 words on list A in the first week, almost all the 30 target words on lists A and B in the second week, and almost all the 45 target words on lists A, B, and C in the third week. Not only was the RMCG successful in recognising almost all the target words in LDTs by the end of each week of learning, they were also successful in producing them in the GFTs. The RMCG reacted to list A faster than to lists B and C, and they also reacted faster to list B than to list C.

The expectation was that learning a new set of 45 target words would result in a dramatic difference in performance between the RMCG and the GRG. The results show that there is a significant difference between the RMCG and the GRG in word recognition of the three lists. The GRG recognised more than half the target words in list A. In the second week of learning the GRG recognised a considerable number of target words in lists A and B. The addition of list C did not affect the GRG performance and they were able to recognise 30.48 out of 45 words from the three lists. As for the GRG RT for each of the lists, the GRG reacted faster to list A than to lists B and C. Also, the GRG reacted slightly faster to list C than to list B.

It was also expected that a dramatic difference in the production of target words would be observed between the GRG and the RMCG. The results show that there was a significant difference between the RMCG and the GRG in word production of the three lists. The GRG production of words from lists A, B, and C declined when there was a gap between exposure and the coming weekly immediate post-test. Also, as expected, the CG came last in all the weekly immediate post-tests. Moreover, the results show that there was a significant difference between the CG and both the RMCG and the GRG in word recognition of the three lists. The CG failed to produce target words from the three lists. The CG reacted to list A faster than to list B, and reacted to list C faster than to list B in LDT 4.

The RMCG performance on the three lists

There are three main reasons the RMCG had no difficulty in acquiring almost all the 15 target words in list A in the first weekly immediate post-test. The RMCG rehearsed list A words using the *Leitner system* in four sessions (over four subsequent days), and this allowed *retrieval practice. Retrieval practice* was more effective when taking into consideration time and the number of spaced learning sessions needed to fulfil the purpose of distributing learning (Nakata, 2008). Second, the group received the first weekly immediate post-test during the fifth session (on the fifth day) just after learning. Third, list A was the first list introduced and there was no learning burden before memorising it. Furthermore, it was a short list of words (i.e., 15 target words) to learn in five days in comparison to previous studies. For example, the subject Sue, in the Fitzpatrick et al. (2008) study, was required to learn a word list of 400 Arabic words over a period of 20 days which, if divided equally, is approximately 133 words a week.

Like list A, adding list B to the *Leitner system* in the four sessions (on four subsequent days) allowed *retrieval practice.* Furthermore, as for list A, the RMCG had the second weekly immediate post-test immediately after the fifteenth session and retained target words from list B while they were still fresh in their minds. However, the second weekly immediate post-test was more challenging because the RMCG had a higher learning burden and this explained the slightly lower recognition and production of list B compared to list A. It seems that even with a delay from the last exposure to list A, this did not affect the RMCG performance for list A words in LDT 2.

In the third week, the delay of the last exposure to list B, and the longer delay to the last exposure to list A, did not affect the RMCG performance in LDT 3. Like lists A and B, list C words had the same advantage of *retrieval practice* and the same advantage of the immediate testing while target words remain fresh in the mind. However, list C was more challenging for the RMCG because the learning burden was higher (i.e., in addition to list C are lists A and B).

Not only was the RMCG successful in recognising almost all the target words in LDTs by the end of one week of learning, they were also successful in the GFTs. Rehearsal using word cards and the *Leitner system* was beneficial because the RMCG was able to move from the first part of the continuum, which is recognising the form of the words because of frequent exposure or training, until they reach the point where they can produce the words, as suggested by Meara 1990.

In the third LDT there was a noticeable difference between the RT of the three lists in the RMCG performance. The familiarity with the words from list A as it was given to the learners in the early stage of learning (i.e., week one), and for a longer time (i.e., two weeks in the Leitner box), contributed to the faster reaction of the RMCG to list A. Similarly, list B was seen by the RMCG for a longer time than list C (i.e., two weeks in the Leitner box), and the learning burden was lower than when they were introduced to list C. As the RMCG were introduced to list C, just before taking LDT 3, they were slower in their RT to list C words than they were to list A and B words. This was the first time they were tested on list C words and shows that they were careful in responding to stimuli from list C which limits the possibility of having a high AS by chance.

The GRG performance on the three lists

The high numbers of recognition of the three lists confirms the practicality of the design of graded readers according to the suggestion of Laufer (1989) and Nation (2001), which is that learner's need 95% coverage of the words in a text to understand it. Also, this successful word recognition of the GRG in the first week confirms the efficiency of repeating target words ten times, as suggested by Webb (2007a, 2007b).

Generally, the GRG higher numbers of recognition of a list occurred when the group was tested for the first time. The recent exposure to a list in the graded readers might have contributed to better recognition as the list might have been fresh in the GRG's minds. The gap between the exposure to a

list and the time of the test does not affect the GRG performance dramatically. It seems that the GRG retained target words from list A, even with a ten day gap, and target words from list B, even with five days' gap between learning them to the time of the weekly immediate post-test.

It was harder for the GRG to produce more than half of the 45 target words as it was expected that incidental learning from the graded readers can occur at the recognition level, rather than at the production level of the continuum, as suggested by Meara (1990). Unlike intentional learning, incidental learning from graded readers needs more time and the results usually show that, even by applying the conditions needed for graded readers to be successful, there remains a chance that the GRG would not pay attention to the unknown words and would try to pass over them while reading. Even when paying attention to the target words this does not ensure learning them. The learner might not pay attention to all repetitions, and may recognise the form without knowing the meaning.

The GRG reacted to the three lists with very similar RTs. This shows that despite the fact that the GRG received the three lists at three different times, and despite the gap between the presentations of the three lists and taking LDT 4 two weeks after the learning period, they reacted similarly to each list.

The CG performance on the three lists

The CG was not exposed to any of the target words in any way, their performance provides evidence of how the word cards in conjunction with the *Leitner system* on the one hand, and the graded readers on the other hand, were effective in participants recognizing and producing target words. Of course, there is the possibility of guessing while taking the LDT, and this could be the cause of false scoring for some of the target words in the LDT. However, even if the correct responses in the LDT occurred by chance, there remains a robustly significant difference between the CG and the RMCG, and between the CG and the GRG. The CG failed to produce almost any of the target words and this shows that the learning performance of the RMCG and the GRG can be attributed to the use of word cards, in conjunction with the *Leitner system*, and graded readers.

The performance of the three groups in the three lists

The RMCG outperformed the GRG and the CG. Similar to the weekly tests results, the post-test shows that the RMCG had the highest LDT AS and GFT scores, followed by the GRG and the CG. The RMCG recognised 34.04 of the 45 target words two weeks after the end of the learning phase. This shows that word cards in conjunction with the *Leitner system* were effective and helped the RMCG to retain and recognise more than half the number of target words, even with no recent exposure to them. The RMCG was also able to produce 25.44 out of 45 target words. This result shows that intentional learning from word cards, in conjunction with the *Leitner system,* could be effective for long term retention.

The GRG was also able to recognise a good number of target words after two weeks from learning (i.e., 31.46 out of 45). The GRG successfully produced fewer target words than the RMCG (i.e., 15.16 out of 45) after two weeks from learning. Although the number of target words produced by the GRG were very few in comparison to the number of words produced by the RMCG, it was not expected that the GRG would produce target words.

The CG recognition was not very high, but it was higher than their production in the weekly immediate post-tests. In the third weekly test the CG recognised 17.28 target words, but in the post-test the CG recognised 20.36 target words out of 45. This is discussed further in section 1.3.5. The CG produced .04 target words in the GFT taken two weeks after the last weekly test, which is a very small number.

Appendix 6. Main Study Results
F. Target words ordered from the most recognized and produced to the least

Target words ordered from the most recognized and produced by the RMCG and the GRG to the least recognized and produced by the RMCG and the GRG

word	RMCG LDT AS	GRG LDT AS	RMCG GFT	GRG GFT
	average 77%	average 70%	average 63%	average 44%
new-fangled /njuːˈfaŋgəld/	92%	85%	89%	61%
shrewd /ʃruːd/	92%	80%	89%	53%
trespass /ˈtrɛspəs	91%	83%	88%	56%
ingenious /ɪnˈdʒiːnɪəs/	89%	82%	84%	58%
proprietorship /prəˈprʌɪətəʃɪp/	87%	76%	77%	51%
concur /kənˈkəː/	85%	74%	67%	45%
banish /ˈbanɪʃ/	83%	75%	94%	69%
bonny /ˈbɒni/	80%	76%	84%	69%
vicissitude /vɪˈsɪsɪtjuːd/	86%	76%	85%	41%
contentious /kənˈtɛnʃəs/	85%	71%	68%	43%
obstinate /ˈɒbstɪnət/	84%	73%	71%	43%
contrive /kənˈtrʌɪv/	85%	69%	73%	54%
colossal /kəˈlɒsəl/	85%	68%	75%	47%
elucidate /ɪˈljuːsɪdeɪt/	83%	75%	73%	25%

stupendous /stjuːˈpɛndəs/	80%	57%	72%	63%
irksome /ˈəːksəm/	79%	65%	73%	61%
perceptive /pəˈsɛptɪv/	89%	74%	61%	41%
idiosyncratic /ˌɪdɪəʊsɪŋˈkratɪk/	83%	72%	55%	36%
adroit /əˈdrɔɪt/	77%	63%	67%	59%
sluggishly /ˈslʌgɪʃəli/	76%	80%	58%	52%
brittle /ˈbrɪtəl/	74%	73%	45%	53%
misgiving /mɪsˈgɪvɪŋ/	73%	53%	75%	49%
astound /əˈstaʊnd/	72%	74%	66%	40%
folly /ˈfɒli/	84%	70%	62%	44%
pertinent /ˈpəːtɪnənt/	83%	62%	61%	32%
postulation /pɒstjʊˈleɪʃən/	82%	70%	58%	27%
lucid /ˈluːsɪd/	80%	70%	56%	36%
rambunctious /ramˈbʌŋkʃəs/	80%	66%	38%	26%
mystification /mɪstɪfɪˈkeɪʃən/	79%	63%	55%	21%
ebullient /ɪˈbʌljənt/	79%	61%	63%	41%
gargantuan /gɑːˈgantjʊən/	76%	71%	61%	36%

acrid /ˈakrɪd/	76%	68%	73%	39%
detestable /dɪˈtɛstəbəl/	76%	72%	55%	35%
apprehensive /aprɪˈhɛnsɪv/	73%	71%	57%	31%
tedious /ˈtiːdɪəs/	72%	68%	60%	64%
veritably /ˈvɛrɪtəbli/	66%	78%	62%	36%
feeble /ˈfiːbəl/	62%	80%	50%	40%
cogent /ˈkəʊdʒənt/	58%	60%	50%	52%
perseverance /pəːsɪˈvɪərəns/	76%	62%	44%	20%
expeditious /ˌɛkspɪˈdɪʃəs/	76%	68%	36%	32%
afflicted /əˈflɪktəd/	60%	66%	34%	34%
unequivocally /ˌʌnɪˈkwɪvəkəli/	72%	68%	63%	43%
grievous /ˈgriːvəs/	68%	68%	48%	30%
petrify /ˈpɛtrɪfʌɪ/	64%	66%	44%	26%
toil /tɔɪl/	64%	66%	64%	48%

Appendix 6. Main Study Results
G. Target words and derivational morphemes

1. Target words with derivational morphemes and percentage of participant recognition in LDT and production in GFT

Target words with suffixes and percentage of participant recognition in LDT and production in GFT

Level of frequency	Derivational ending	Target word	RMCG LDT AS	GRG LDT AS	CG LDT AS	RMCG GFT	GRG GFT
			average 77%	average 70%	average 41%	average 63%	average 44%
Low frequency	*-some*	irksome	79%	65%	44%	73%	61%
	-itude	vicissitude	86%	76%	39%	85%	41%
	-ship	proprietorship	87%	76%	46%	77%	51%
	-ify	petrify	64%	66%	40%	44%	26%
	-ate (adj)	obstinate	84%	73%	40%	71%	43%
	-ance	perseverance	76%	62%	50%	44%	20%
	-ent	pertinent	76%	62%	35%	61%	32%
		ebullient	79%	61%	33%	63%	41%
		cogent	58%	60%	38%	50%	52%
	-an	gargantuan	76%	71%	39%	61%	36%
Mid frequency	*-ive*	perceptive	89%	74%	40%	61%	41%
		apprehensive	73%	71%	36%	57%	31%
		contrive	85%	69%	43%	73%	54%
	-ous	ingenious	89%	82%	43%	84%	58%
		stupendous	80%	57%	53%	72%	63%
		grievous	68%	68%	38%	48%	30%
		tedious	72%	68%	42%	60%	64%
		expeditious	76%	68%	30%	36%	32%
		rambunctious	80%	66%	38%	38%	26%
		contentious	85%	71%	32%	68%	43%
	-ation	mystification	79%	63%	46%	55%	21%
	-able	detestable	76%	72%	37%	55%	35%
	-ate (V)	elucidate	83%	75%	30%	73%	25%
	-ic	idiosyncratic	83%	72%	39%	55%	36%
	-al	colossal	85%	68%	41%	75%	47%
	-ion	postulation	82%	70%	34%	58%	27%
High frequency	*-ly*	unequivocally	72%	68%	43%	63%	43%
		veritably	66%	78%	52%	62%	36%
		sluggishly	76%	80%	50%	58%	52%
	-ing	misgiving	73%	53%	41%	75%	49%
	-ed	new-fangled	92%	85%	43%	89%	61%
		afflicted	60%	66%	30%	34%	34%

2. Frequency of Suffix Categories in Spoken British English (from the BNC):

Suffix	Part of Speech (PoS) of derivative	Frequency level	Occurrences in Target List	Type Frequency (target PoS)	Type Frequency (other PoS)	Token Frequency (target PoS)	Token Frequency (other PoS)
-some	(adj)	Low frequency	1	14	5 (n)	90	12 (n)
-itude	(n)		1	15		612	
-ship	(n)		1	54		2,366	
-ify	(v)		1	54		1,846	
-ate	(adj form)		1	58	49 (n)	3,645	1,549
-ance	(n)		1	103		6,262	
-ent	(adj)		3	105	44 (n)	9,758	4,817 (n)
-an	(adj)		1	141	98 (n)	3,654	1,119 (n)
-ive	(adj)	Mid frequency	2	209	57 (n)	7,193	2,995 (n)
-ous	(adj)		7	261		8,116	
-ation	(n)		1	267		12,570	
-able	(adj)		1	304	16 (n)	6,381	370 (n)
-ate	(verb form)		1	314		10,217	
-ic	(adj)		1	391	81 (n)	7,500	3,296 (n)
-al	(adj)		1	558	181 (n)	37,583	9,853 (n)
-ion	(n)		1	600		52,016	
-ly	(adv)	High frequency	3	1,140	66 (adj)	111,045	8,893 (adj)
-ing	(n)		1	1,425	940 (adj)	32,543	13,665 (adj)
-ed	(adj)		2	1,593	24 (n)	29,633	120 (n)

Data from http://morphoquantics.co.uk

Appendix 6. Main Study Results
H. Word length and percentage of participant recognition in LDT and production in GFT

Percentage of participant recognition in LDT and production in GFT and word length and number of syllables.

Number of syllables	**word**	**RMCG LDT AS**	**GRG LDT AS**	**CG LDT AS**	**RMCG GFT**	**GRG GFT**
		average 77%	average 70%	average 41%	average 63%	average 44%
Six syllables	id.i.o.syn.crat.ic	83%	72%	39%	55%	36%
	un.e.quiv.o.cal.ly	72%	68%	43%	63%	43%
Five syllables	pro.pri.e.tor.ship	87%	76%	46%	77%	51%
	mys.ti.fi.ca.tion	79%	63%	46%	55%	21%
Four syllables	vi.cis.si.tude	86%	76%	39%	85%	41%
	pos.tu.la.tions	82%	70%	34%	58%	27%
	de.test.a.ble	76%	72%	37%	55%	35%
	e.lu.ci.date	83%	75%	30%	73%	25%
	ap.pre.hen.sive	73%	71%	36%	57%	31%
	ex.pe.di.tious	76%	68%	30%	36%	32%
	per.se.ver.ance	76%	62%	50%	44%	20%
	ver.i.ta.bly	66%	78%	52%	62%	36%
	gar.gan.tu.an	76%	71%	39%	61%	36%
Three syllables	con.ten.tious	85%	71%	32%	68%	43%
	new.fan.gled	92%	85%	43%	89%	61%
	per.cep.tive	89%	74%	40%	61%	41%
	in.gen.ious	89%	82%	43%	84%	58%
	per.ti.nent	83%	62%	35%	61%	32%
	a.dr.oit	77%	63%	43%	67%	59%
	stu.pen.dous	80%	57%	53%	72%	63%
	co.los.sal	85%	68%	41%	75%	47%
	ob.sti.nate	84%	73%	40%	71%	43%
	e.bul.lient	79%	61%	33%	63%	41%
	mis.giv.ing	73%	53%	41%	75%	49%
	af.flict.ed	60%	66%	30%	34%	34%
	ram.bunc.tious	80%	66%	38%	38%	26%
	slug.gish.ly	76%	80%	50%	58%	52%

	pet.ri.fy	64%	66%	40%	44%	26%
	te.di.ous	72%	68%	42%	60%	64%
Two syllables	con.cur	85%	74%	41%	67%	45%
	con.trive	85%	69%	43%	73%	54%
	ban.ish	83%	75%	48%	94%	69%
	tres.pass	91%	83%	42%	88%	56%
	as.tound	72%	74%	36%	66%	40%
	ac.rid	76%	68%	36%	73%	39%
	irk.some	79%	65%	44%	73%	61%
	co.gent	58%	60%	38%	50%	52%
	griev.ous	68%	68%	38%	48%	30%
	bon.ny	80%	76%	44%	84%	69%
	brit.tle	74%	73%	45%	45%	53%
	fol.ly	84%	70%	40%	62%	44%
	feeb.le	62%	80%	48%	50%	40%
	lu.cid	80%	70%	48%	56%	36%
One syllable	shrewd	92%	80%	47%	89%	53%
	toil	64%	66%	44%	64%	48%
mean of participant percentage in each test		78%	70%	41%	64%	44%

Appendix 6. Main Study Results

I. Target words with difficult sounds in Arabic and percentage of participant recognition in LDT and production in GFT

Target words with difficult sounds in Arabic and percentage of participant recognition in LDT and production in GFT

Word and phonetic prescription	Sound that might cause difficulty	RMCG LDT AS	GRG LDT AS	CG LDT AS	RMCG GFT	GRG GFT
		average 77%	average 70%	average 41%	average 63%	average 44%
List A						
vicissitude /vɪˈsɪsɪtjuːd/	/v/	86%	76%	39%	85%	41%
proprietorship /prəˈpraɪətəʃɪp/	/p/	87%	76%	46%	77%	51%
postulation /ˈpɒstjuleɪeɪʃn/	/p/	82%	70%	34%	58%	27%
new-fangled /ˌnjuːˈfæŋgld/	/g/	92%	85%	43%	89%	61%
perceptive /pəˈseptɪv/	/p/ /v/	89%	74%	40%	61%	41%
trespass /ˈtrespəs/	/p/	91%	83%	42%	88%	56%
contrive /kənˈtraɪv/	/v/ /	85%	69%	43%	73%	54%
pertinent /ˈpɜːtɪnənt/	/p/ /r/	83%	62%	35%	61%	32%
List B						
stupendous /stjuːˈpendəs/	/p/	80%	57%	53%	72%	63%
misgiving /ˌmɪsˈgɪvɪŋ/	/v/	73%	53%	41%	75%	49%

gargantuan /gɑːˈgæntʃuən/	/g/	76%	71%	39%	61%	36%
apprehensive /ˌæprɪˈhensɪv/	/p/	73%	71%	36%	57%	31%
unequivocally /ˌʌnɪˈkwɪvəkəli/	/v/	72%	68%	43%	63%	43%
List C						
grievous /ˈgriːvəs	/g/	68%	68%	38%	48%	30%
veritably /ˈverɪtəbli/	/v/	66%	78%	52%	62%	36%
sluggishly /ˈslʌgɪʃli/	/g/	76%	80%	50%	58%	52%
petrify /ˈpetrɪfaɪ/	/p/	64%	66%	40%	44%	26%
perseverance /ˌpɜːsɪˈvɪərəns/	/p/	76%	62%	50%	44%	20%
Consonant cluster						
proprietorship /prəˈpraɪətəʃɪp/	/pr/	87%	76%	46%	77%	51%
Mirror shape letters						
detestable /dɪˈtestəbl/	/d/ and /b/ mirror shape	76%	72%	37%	55%	35%

Appendix 6. Main Study Results
J. Long and short vowels in target words

Target words and vowels ordered from the most recognized and produced by the RMCG and the GRG to the least recognized and produced by the RMCG and the GRG

word		RMCG LDT AS	GRG LDT AS	RMCG GFT	GRG GFT
		average 77%	average 70%	average 63%	average 44%
new-fangled /njuːˈfaŋgəld/	a, uː	92%	85%	89%	61%
shrewd /ʃruːd/	uː	92%	80%	89%	53%
trespass /ˈtrɛspəs	ɛ	91%	83%	88%	56%
ingenious /ɪnˈdʒiːnɪəs/	ɪ, iː	89%	82%	84%	58%
proprietorship /prəˈprʌɪətəʃɪp/	ʌ , ɪə	87%	76%	77%	51%
concur /kənˈkəː/	əː	85%	74%	67%	45%
banish /ˈbanɪʃ/	a, ɪ	83%	75%	94%	69%
bonny /ˈbɒni/	ɒ, i	80%	76%	84%	69%
vicissitude /vɪˈsɪsɪtjuːd/	ɪ, ɪ , ɪ uː	86%	76%	85%	41%
contentious /kənˈtɛnʃəs/	ɛ	85%	71%	68%	43%
obstinate /ˈɒbstɪnət/	ɒ, ɪ	84%	73%	71%	43%
contrive /kənˈtrʌɪv/	ʌɪ	85%	69%	73%	54%
colossal /kəˈlɒsəl/	ɒ	85%	68%	75%	47%
elucidate /ɪˈljuːsɪdeɪt/	ɪ, uː, ɪ	83%	75%	73%	25%

stupendous /stjuːˈpɛndəs/	uː, ɛ	80%	57%	72%	63%
irksome /ˈəːksəm/	əː	79%	65%	73%	61%
perceptive /pəˈsɛptɪv/	ɛ , ɪ	89%	74%	61%	41%
idiosyncratic /ˌɪdɪəʊsɪŋˈkratɪk/	ɪ, ɪ, ɪ, əʊ	83%	72%	55%	36%
adroit /əˈdrɔɪt/	ɔɪ	77%	63%	67%	59%
sluggishly /ˈslʌgɪʃəli/	ʌ, ɪ, i	76%	80%	58%	52%
brittle /ˈbrɪtəl/	ɪ	74%	73%	45%	53%
misgiving /mɪsˈgɪvɪŋ/	ɪ, ɪ, ɪ	73%	53%	75%	49%
astound /əˈstaʊnd/	aʊ	72%	74%	66%	40%
folly /ˈfɒli/	ɒ , i	84%	70%	62%	44%
pertinent /ˈpəːtɪnənt/	əː, ɪ	83%	62%	61%	32%
postulation /pɒstjʊˈleɪʃən/	ɒ, ʊ, eɪ	82%	70%	58%	27%
lucid /ˈluːsɪd/	uː, ɪ	80%	70%	56%	36%
rambunctious /ramˈbʌŋkʃəs/	a, ʌ	80%	66%	38%	26%
mystification /mɪstɪfɪˈkeɪʃən/	ɪ , ɪ, ɪ	79%	63%	55%	21%
ebullient /ɪˈbʌljənt/	ɪ, ʌ	79%	61%	63%	41%
gargantuan /gɑːˈgantjʊən/	ɑː, a, ʊə	76%	71%	61%	36%

acrid /ˈakrɪd/	a, ɪ	76%	68%	73%	39%
detestable /dɪˈtɛstəbəl/	ɪ, ɛ	76%	72%	55%	35%
apprehensive /aprɪˈhɛnsɪv/	ɪ, ɛ, ɪ	73%	71%	57%	31%
tedious /ˈtiːdɪəs/	iː, ɪə	72%	68%	60%	64%
veritably /ˈvɛrɪtəbli/	ɛ, ɪ, i	66%	78%	62%	36%
feeble /ˈfiːbəl/	iː	62%	80%	50%	40%
cogent /ˈkəʊdʒənt/	əʊ	58%	60%	50%	52%
perseverance /pəːsɪˈvɪərəns/	əː, ɪ, ɪə	76%	62%	44%	20%
expeditious /ˌɛkspɪˈdɪʃəs/	ɛ, ɪ, ɪ	76%	68%	36%	32%
afflicted /əˈflɪktəd/	ɪ	60%	66%	34%	34%
unequivocally /ˌʌnɪˈkwɪvəkəli/	ʌ, ɪ, ɪ, ɪ	72%	68%	63%	43%
grievous /ˈgriːvəs/	iː	68%	68%	48%	30%
petrify /ˈpɛtrɪfʌɪ/	ɛ, ɪ, ʌɪ	64%	66%	44%	26%
toil /tɔɪl/	ɔɪ	64%	66%	64%	48%

Appendix 6. Main Study Results
K. A comparison of RMCG, GRG and CG RT of each target word in list A in the four LDTs.

1- RMCG RT

Mean of LDT RTs for each word in list A for the RMCG.

List A	LDT RT 1	LDT RT 2	LDT RT 3	LDT RT 4	mean of LDT RT for each word
vicissitude	2147.35	1912.17	1853.67	2445.37	2089.64
proprietorship	1909.75	2321.50	1718.59	2463.79	2103.41
postulation	3158.28	2225.38	1703.36	2478.10	2391.28
contentious	2404.70	2154.65	1810.41	2125.38	2123.78
ingenious	2197.74	1849.43	1686.18	1591.43	1831.20
new-fangled	2440.58	2262.76	1812.58	1671.23	2046.79
perceptive	2621.43	3342.55	2093.62	2013.76	2517.84
concur	2054.14	1830.10	1713.68	1737.40	1833.83
elucidate	2419.79	2220.67	1681.90	2414.14	2184.13
shrewd	1862.39	1673.96	1388.26	2057.91	1745.63
trespass	1882.14	2178.48	1724.09	1670.30	1863.75
contrive	2346.09	2212.29	1614.19	1974.48	2036.76
detestable	3273.17	2750.90	2016.15	2149.30	2547.38
banish	2296.00	2178.55	1392.73	2314.14	2045.35
pertinent	2656.94	2227.05	1801.15	2339.68	2256.21
mean of LDT RT for list A in each test	2378.03	2222.70	1734.04	2096.43	2107.79

2- GRG RT

Mean of LDT RTs for each word in list A for the GRG.

List A	LDT RT 1	LDT RT 2	LDT RT 3	LDT RT 4	mean of LDT RT for each word
vicissitude	2607.11	2082.00	1720.36	1886.61	2074.02
proprietorship	2281.37	2534.33	2146.67	1975.05	2234.36
postulation	2694.60	2227.88	1841.21	1930.76	2173.62
contentious	2515.57	2769.41	1695.06	1393.60	2093.41
ingenious	2201.05	2160.00	1390.35	1360.23	1777.91
new-fangled	2508.52	2914.11	2219.16	1361.73	2250.88
perceptive	3048.58	2608.80	2523.57	1888.94	2517.47
concur	2287.10	2042.29	1749.47	1668.35	1936.80

elucidate	3659.95	3062.76	1905.65	1737.72	2591.52
shrewd	2250.05	1702.61	1481.26	1395.28	1707.30
trespass	3252.29	2649.38	2346.70	1902.79	2537.79
contrive	3274.75	2204.17	1800.35	1679.67	2239.73
detestable	3235.53	2756.37	2690.35	2377.71	2764.99
banish	2764.59	2156.24	1597.47	1507.65	2006.49
pertinent	2771.83	3423.54	1909.43	1767.00	2467.95
mean of LDT RT for list A in each test	2756.86	2486.26	1934.47	1722.21	2224.94

3. CG RT

Mean of LDT RTs for each word in list A for the CG.

List A	LDT RT 1	LDT RT 2	LDT RT 3	LDT RT 4	mean of LDT RT for each word
vicissitude	2775.14	1897.67	1686.82	1220.71	1895.09
proprietorship	2750.30	3073.91	1394.60	1901.71	2280.13
postulation	3803.50	2521.75	1741.50	1543.64	2402.60
contentious	1853.60	1946.00	1476.63	1584.91	1715.28
ingenious	3325.00	2134.42	1573.60	1810.46	2210.87
new-fangled	2689.63	2935.09	2051.27	1648.15	2331.04
perceptive	3047.75	2169.44	2225.43	1711.00	2288.41
concur	2129.38	1465.00	2142.63	1378.55	1778.89
elucidate	3476.57	2741.71	1680.63	1687.75	2396.67
shrewd	1951.36	2152.70	1645.27	1593.17	1835.63
trespass	4158.13	2455.00	2048.90	1480.69	2535.68
contrive	3738.82	3379.45	1730.64	1311.31	2540.05
detestable	2621.38	2790.38	1906.44	1276.00	2148.55
banish	3289.13	1240.45	1474.09	1290.40	1823.52
pertinent	3935.25	2764.88	2635.50	1470.70	2701.58
mean of LDT RT for list A in each test	3036.33	2377.86	1827.60	1527.28	2192.26

Appendix 6. Main Study Results
L. A comparison of RMCG, GRG and CG RT of each target word in list B in the three LDTs.

1. RMCG RT

Mean of LDT RTs for each word in list B for the RMCG.

List B	LDT RT 2	LDT RT 3	LDT RT 4	mean of LDT RT for each word
adroit	2304.00	2231.85	1637.44	2057.76
stupendous	2174.79	2196.14	1942.25	2104.39
misgiving	2407.25	2211.00	1661.44	2093.23
gargantuan	2214.52	1992.41	2888.18	2365.04
irksome	1708.95	1793.48	2094.44	1865.62
mystification	2636.06	2155.09	1367.83	2052.99
ebullient	2225.47	2508.82	1818.21	2184.17
idiosyncratic	2630.39	2523.67	1295.86	2149.97
acrid	2119.24	1744.04	1596.19	1819.82
obstinate	1958.29	2070.13	1638.05	1888.82
colossal	2141.68	2244.39	1727.62	2037.90
bonny	1788.00	1723.33	2531.26	2014.20
apprehensive	2611.80	2325.10	2894.27	2610.39
brittle	2283.85	2078.36	1607.06	1989.76
unequivocally	2781.00	2140.40	2369.71	2430.37
mean of LDT RT for list B in each test	2265.69	2129.21	1937.99	2110.96

2. GRG RT

Mean of LDT RTs for each word in list B for the GRG.

List B	LDT RT 2	LDT RT 3	LDT RT 4	mean of LDT RT for each word
adroit	2473.88	2825.40	1363.29	2220.86
stupendous	2244.76	2485.07	1952.27	2227.37
misgiving	2274.92	3435.14	2148.71	2619.59
gargantuan	2371.05	3203.83	2469.89	2681.59
irksome	2062.59	2158.93	2132.63	2118.05
mystification	2117.64	2264.50	2018.18	2133.44
ebullient	1899.45	2177.64	1723.07	1933.39
idiosyncratic	2444.00	2535.31	1659.94	2213.08
acrid	2670.88	1791.14	1444.84	1968.95

obstinate	2168.42	2308.41	1464.13	1980.32
colossal	2111.53	2101.33	2035.35	2082.74
bonny	1962.37	2244.30	1767.55	1991.41
apprehensive	2926.79	2082.50	2266.35	2425.21
brittle	2808.40	2775.89	1281.67	2288.65
unequivocally	3044.64	3481.17	1548.88	2691.56
mean of LDT RT for list B in each test	2372.09	2524.70	1818.45	2238.414

3. CG RT

Mean of LDT RTs for each word in list B for the CG.

List B	LDT RT 2	LDT ART 3	LDT RT 4	mean of LDT RT for each word
adroit	2474.50	1477.70	2575.67	2175.96
stupendous	2289.38	1036.36	2240.35	1855.37
misgiving	2100.00	1576.73	2033.43	1903.39
gargantuan	2563.82	1308.11	1897.17	1923.03
irksome	1964.56	2216.90	1475.62	1885.69
mystification	1529.73	1861.20	1845.36	1745.43
ebullient	2393.25	2119.29	1884.33	2132.29
idiosyncratic	2062.38	2719.78	1654.25	2145.47
acrid	1896.25	1965.10	1326.31	1729.22
obstinate	1538.36	2152.11	1636.56	1775.68
colossal	2160.00	1790.67	2235.80	2062.16
bonny	1624.91	1657.67	2596.50	1959.69
apprehensive	3021.71	1367.63	1443.40	1944.25
brittle	2488.30	1285.38	2257.82	2010.50
unequivocally	1996.33	1934.00	1398.27	1776.20
mean of LDT RT for list B in each test	2140.23	1764.57	1900.05	1934.95

Appendix 6. Main Study Results
M. A comparison of RMCG, GRG and CG RT of each target word in list C in the last two LDTs

1. RMCG RT

Mean of LDT RTs for each word in list C for the RMCG.

List C	LDT RT 3	LDT RT 4	mean of LDT RT for each word
grievous	2394.71	1676.19	2035.45
cogent	2652.89	1955.43	2304.16
toil	1977.76	2656.81	2317.29
tedious	2416.26	2576.70	2496.48
expeditious	2854.29	3336.90	3095.60
astound	2112.93	2093.47	2103.20
veritably	2372.56	2004.44	2188.50
feeble	1970.06	1757.13	1863.59
afflicted	3461.37	1607.20	2534.28
sluggishly	2954.91	2197.70	2576.30
petrify	3336.11	1817.44	2576.77
lucid	2372.94	2556.74	2464.84
perseverance	2061.80	2046.06	2053.93
folly	2443.52	2226.50	2335.01
rambunctious	2696.47	2122.22	2409.34
mean of LDT RT for list C in each test	2538.57	2175.40	2356.98

2. GRG RT

Mean of the LDT RTs for each word in list C for the GRG.

List C	LDT RT 3	LDT RT 4	mean of LDT RT for each word
grievous	2826.50	2073.94	2450.22
cogent	2199.56	1713.63	1956.59
toil	2296.57	1825.56	2061.06
tedious	2625.74	1828.05	2226.89
expeditious	2475.13	2020.11	2247.62
astound	2151.38	1681.76	1916.57
veritably	2335.07	1904.28	2119.67
feeble	2099.93	1653.15	1876.54
afflicted	1810.00	2815.07	2312.53
sluggishly	2474.42	1818.67	2146.54
petrify	2486.20	1787.05	2136.63
lucid	2046.05	1334.65	1690.35

perseverance	2453.00	1679.53	2066.26
folly	2733.76	1758.15	2245.96
rambunctious	2989.33	1862.75	2426.04
mean of LDT RT for list C in each test	2400.18	1850.42	2125.298

3. CG RT

Mean of LDT RTs for each word in list C for the CG.

List C	LDT RT 3	LDT RT 4	mean of LDT RT for each word
grievous	1690.44	1678.92	1684.68
cogent	2603.00	1631.58	2117.29
toil	1731.55	1516.00	1623.77
tedious	2155.22	1469.30	1812.26
expeditious	2128.00	1826.63	1977.31
astound	2160.86	1845.30	2003.08
veritably	2267.60	1644.33	1955.97
feeble	1843.33	1597.85	1720.59
afflicted	2013.00	1950.86	1981.93
sluggishly	1190.75	1612.88	1401.82
petrify	2016.60	1571.00	1793.80
lucid	1385.09	1296.14	1340.62
perseverance	1481.29	1773.25	1627.27
folly	1637.82	1665.29	1651.55
rambunctious	1641.43	1916.63	1779.03
mean of LDT RT for list C in each test	1863.07	1666.40	1764.73

Appendix 6. Main Study Results
N. A comparison of the RMCG, GRG and CG recognition of each target word in list A in the four LDTs

1. RMCG AS

The percentage of number of participants in RMCG who recognised target words in list A in LDT 1, LDT 2, LDT 3 and LDT 4.

List A	test				Mean of the participant recognition for each word
	LDT AS 1	LDT AS 2	LDT AS 3	LDT AS 4	
vicissitude	88%	92%	88%	76%	86%
proprietorship	84%	92%	96%	76%	87%
postulation	76%	84%	84%	84%	82%
contentious	84%	84%	88%	84%	85%
ingenious	92%	92%	88%	84%	89%
new-fangled	100%	88%	92%	88%	92%
perceptive	96%	84%	92%	84%	89%
concur	88%	88%	84%	80%	85%
elucidate	80%	88%	80%	84%	83%
shrewd	96%	92%	88%	92%	92%
trespass	92%	100%	80%	92%	91%
contrive	92%	88%	76%	84%	85%
detestable	76%	88%	84%	80%	82%
banish	92%	88%	68%	84%	83%
pertinent	76%	84%	84%	88%	83%
Mean of the participant recognition of list A in each test	87%	89%	85%	84%	

2. GRG AS

The percentage of GRG participants who recognised target words in list A in LDT 1, LDT 2, LDT 3 and LDT 4.

List A	test				Mean of the participant recognition for each word
	LDT AS 1	LDT AS 2	LDT AS 3	LDT AS 4	
vicissitude	72%	70%	88%	72%	76%
proprietorship	80%	72%	76%	76%	76%
postulation	80%	68%	64%	68%	70%
contentious	84%	68%	72%	60%	71%
ingenious	88%	84%	68%	88%	82%
new-fangled	88%	84%	80%	88%	85%

perceptive	80%	64%	80%	72%	74%
concur	88%	72%	68%	68%	74%
elucidate	84%	76%	68%	72%	75%
shrewd	84%	88%	76%	72%	80%
trespass	88%	96%	70%	76%	83%
contrive	80%	72%	64%	60%	69%
detestable	84%	76%	60%	68%	72%
banish	92%	84%	56%	68%	75%
pertinent	76%	56%	60%	56%	62%
Mean of the participant recognition of list A in each test	83%	75%	70%	71%	

3. CG AS

The percentage of CG participants who recognised target words in list A in LDT 1, LDT 2, LDT 3 and LDT 4.

List A	tests				Mean of the participant recognition for each word
	LDT AS 1	LDT AS 2	LDT AS 3	LDT AS 4	
vicissitude	28%	36%	40%	52%	39%
proprietorship	40%	48%	44%	52%	46%
postulation	32%	36%	28%	40%	34%
contentious	20%	36%	32%	40%	32%
ingenious	36%	48%	36%	52%	43%
new-fangled	32%	44%	48%	48%	43%
perceptive	32%	40%	48%	40%	40%
concur	36%	36%	44%	48%	41%
elucidate	28%	28%	40%	24%	30%
shrewd	44%	44%	48%	52%	47%
trespass	32%	44%	40%	52%	42%
contrive	44%	44%	36%	48%	43%
detestable	36%	32%	44%	36%	37%
banish	32%	48%	48%	64%	48%
pertinent	32%	32%	36%	40%	35%
Mean of participant recognition of list A in each test	34%	40%	41%	46%	

Appendix 6. Main Study Results
O. A comparison of RMCG, GRG and CG recognition of each target word in list B in the three LDTs

1. RMCG AS

The percentage of RMCG participants who recognised target words in list B in LDT 2, LDT 3 and LDT 4.

List B	test			Mean of the participant recognition for each word
	LDT AS 2	LDT AS 3	LDT AS 4	
adroit	72%	88%	72%	77%
stupendous	80%	84%	76%	80%
misgiving	64%	84%	72%	73%
gargantuan	88%	72%	68%	76%
irksome	88%	84%	64%	79%
mystification	76%	88%	72%	79%
ebullient	80%	80%	76%	79%
idiosyncratic	92%	72%	84%	83%
acrid	84%	80%	64%	76%
obstinate	88%	84%	80%	84%
colossal	92%	80%	84%	85%
bonny	84%	64%	92%	80%
apprehensive	84%	76%	60%	73%
brittle	84%	70%	68%	74%
unequivocally	76%	72%	68%	72%
Mean of the participant recognition of list B in each test	82%	79%	73%	

2. GRG AS

The percentage of GRG participants who recognised target words in list B in LDT 2, LDT 3 and LDT 4.

List B	test			Mean of the participant recognition for each word
	LDT AS 2	LDT AS 3	LDT AS 4	
adroit	68%	64%	56%	63%
stupendous	68%	44%	60%	57%
misgiving	60%	52%	48%	53%
gargantuan	80%	56%	76%	71%
irksome	72%	60%	64%	65%

mystification	48%	72%	68%	63%
ebullient	60%	76%	48%	61%
idiosyncratic	80%	76%	60%	72%
acrid	76%	68%	60%	68%
obstinate	80%	76%	64%	73%
colossal	84%	52%	68%	68%
bonny	80%	80%	68%	76%
apprehensive	76%	68%	68%	71%
brittle	88%	72%	60%	73%
unequivocally	80%	68%	56%	68%
Mean of the participant recognition of list B in each test	73%	66%	62%	

3. CG AS

The percentage of CG participants who recognised target words in list B in LDT 2, LDT 3 and LDT 4.

List B	test			mean of participant recognition for each word
	LDT AS 2	LDT AS 3	LDT AS 4	
adroit	44%	40%	44%	43%
stupendous	52%	44%	64%	53%
misgiving	40%	32%	52%	41%
gargantuan	44%	36%	36%	39%
irksome	40%	40%	52%	44%
mystification	47%	40%	52%	46%
ebullient	32%	40%	28%	33%
idiosyncratic	32%	40%	44%	39%
acrid	32%	32%	44%	36%
obstinate	48%	36%	36%	40%
colossal	44%	44%	36%	41%
bonny	48%	40%	44%	44%
apprehensive	32%	40%	36%	36%
brittle	44%	48%	44%	45%
unequivocally	52%	36%	40%	43%
Mean of participant recognition of list B in each test	42%	39%	43%	

Appendix 6. Main Study Results
P. A comparison of RMCG, GRG and CG recognition of each target word in list C in the two LDTs

1. RMCG AS

The percentage of RMCG participants who recognised target words in list C in LDT 3 and LDT 4.

list C	test		mean of participant recognition for each word
	LDT AS 3	LDT AS 4	
grievous	72%	64%	68%
cogent	60%	56%	58%
toil	68%	60%	64%
tedious	64%	80%	72%
expeditious	72%	80%	76%
astound	76%	68%	72%
veritably	68%	64%	66%
feeble	64%	60%	62%
afflicted	60%	60%	60%
sluggishly	72%	80%	76%
petrify	56%	72%	64%
lucid	84%	76%	80%
perseverance	84%	68%	76%
folly	88%	80%	84%
rambunctious	88%	72%	80%
Mean of participant recognition of list C in each test	72%	69%	

2. GRG AS

The percentage of GRG participants who recognised target words in list C in LDT 3 and LDT 4.

List C	test		mean of participant recognition for each word
	LDT AS 3	LDT AS 4	
grievous	72%	64%	68%
cogent	56%	64%	60%
toil	60%	72%	66%
tedious	60%	76%	68%
expeditious	64%	72%	68%
astound	80%	68%	74%

veritably	84%	72%	78%
feeble	80%	80%	80%
afflicted	72%	60%	66%
sluggishly	88%	72%	80%
petrify	52%	80%	66%
lucid	60%	80%	70%
perseverance	60%	64%	62%
folly	60%	80%	70%
rambunctious	68%	64%	66%
Mean of participant recognition of list C in each test	68%	71%	

3. CG AS

The percentage of CG participants who recognised target words in list C in LDT 3 and LDT 4.

List C	test		Mean of participant recognition for each word
	LDT AS 3	LDT AS 4	
grievous	36%	40%	38%
cogent	28%	48%	38%
toil	40%	48%	44%
tedious	36%	48%	42%
expeditious	24%	36%	30%
astound	32%	40%	36%
veritably	40%	64%	52%
feeble	48%	48%	48%
afflicted	32%	28%	30%
sluggishly	44%	56%	50%
petrify	32%	48%	40%
lucid	40%	56%	48%
perseverance	48%	52%	50%
folly	32%	48%	40%
rambunctious	40%	36%	38%
Mean of participant recognition of list C in each test	36%	46%	

Appendix 6. Main Study Results
Q. A comparison of RMCG and GRG production of each target word in list A in the two GFT

1. RMCG GFT scores

Percentage of RMCG participant scores for each word in list A in the four GFTs.

words	test				mean of GFT score for each word
	GFT 1	GFT 2	GFT 3	GFT 4	
vicissitude	96%	84%	84%	76%	85%
proprietorship	80%	84%	76%	68%	77%
postulation	68%	68%	56%	40%	58%
contentious	76%	80%	64%	52%	68%
ingenious	88%	88%	84%	76%	84%
new-fangled	96%	96%	76%	88%	89%
perceptive	64%	80%	52%	48%	61%
concur	64%	88%	68%	48%	67%
elucidate	68%	92%	76%	56%	73%
shrewd	92%	92%	92%	80%	89%
trespass	92%	96%	88%	76%	88%
contrive	72%	80%	76%	64%	73%
detestable	48%	72%	56%	44%	55%
banish	100%	96%	92%	88%	94%
pertinent	64%	76%	56%	48%	61%
mean of GFT scores for list A in each test	78%	85%	73%	63%	

2. GRG GFT scores

Percentage of GRG participant scores for each word in list A in the GFTs.

List A	test				mean of GFT score for a word
	GFT 1	GFT 2	GFT 3	GFT 4	
vicissitude	52%	40%	36%	36%	41%
proprietorship	60%	52%	52%	40%	51%
postulation	52%	28%	16%	12%	27%
contentious	52%	44%	44%	32%	43%
ingenious	68%	64%	56%	44%	58%
new-fangled	80%	48%	52%	64%	61%

perceptive	56%	40%	36%	32%	41%
concur	68%	44%	32%	36%	45%
elucidate	36%	24%	24%	16%	25%
shrewd	56%	52%	52%	52%	53%
trespass	72%	52%	56%	44%	56%
contrive	60%	64%	52%	40%	54%
detestable	44%	36%	32%	28%	35%
banish	88%	72%	68%	48%	69%
pertinent	48%	32%	28%	20%	32%
mean of GFT scores for list A in each test	59%	46%	42%	36%	

Appendix 6. Main Study Results
R. A comparison of RMCG and GRG production of each target word in list B in the two GFT

1. RMCG GFT scores

Percentage of RMCG participant scores for each word in list B in the GFTs.

List A	test			mean of GFT score for each word
	GFT 2	GFT 3	GFT 4	
adroit	64%	72%	64%	67%
stupendous	72%	80%	64%	72%
misgiving	80%	76%	68%	75%
gargantuan	56%	72%	56%	61%
irksome	68%	84%	68%	73%
mystification	44%	64%	56%	55%
ebullient	68%	76%	44%	63%
idiosyncratic	48%	64%	52%	55%
acrid	72%	84%	64%	73%
obstinate	76%	80%	56%	71%
colossal	68%	84%	72%	75%
bonny	84%	88%	80%	84%
apprehensive	64%	60%	48%	57%
brittle	48%	48%	40%	45%
unequivocally	52%	76%	60%	63%
mean of GFT scores for list B in each test	64%	74%	59%	

2. GRG GFT scores

Percentage of GRG participant scores for each word in list B in the GFTs.

List B	test			mean of GFT score for a word
	GFT 2	GFT 3	GFT 4	
adroit	80%	56%	40%	59%
stupendous	76%	60%	52%	63%
misgiving	68%	44%	36%	49%
gargantuan	56%	24%	28%	36%
irksome	88%	52%	44%	61%
mystification	40%	12%	12%	21%
ebullient	76%	24%	24%	41%

idiosyncratic	64%	28%	16%	36%
acrid	60%	32%	24%	39%
obstinate	72%	28%	28%	43%
colossal	76%	36%	28%	47%
bonny	80%	64%	64%	69%
apprehensive	60%	16%	16%	31%
brittle	68%	48%	44%	53%
unequivocally	60%	36%	32%	43%
mean of GFT scores for list B in each test	68%	37%	33%	

Appendix 6. Main Study Results
S. A comparison of RMCG and GRG production of each target word in list C in the two GFT

1. RMCG GFT scores

Percentage of RMCG participant scores for each word in list C in the GFTs.

List C	test		mean of GFT score for a word
	GFT 3	GFT 4	
grievous	52%	44%	48%
cogent	56%	44%	50%
toil	64%	64%	64%
tedious	60%	60%	60%
expeditious	36%	36%	36%
astound	76%	56%	66%
veritably	68%	56%	62%
feeble	56%	44%	50%
afflicted	40%	28%	34%
sluggishly	64%	52%	58%
petrify	52%	36%	44%
lucid	64%	48%	56%
perseverance	44%	44%	44%
folly	68%	56%	62%
rambunctious	44%	32%	38%
mean of GFT scores for list C in each test	56%	46%	

2. GRG GFT scores

Percentage of GRG participant scores for each word in list C in the GFTs.

List C	test		mean of GFT score for a word
	GFT 3	GFT 4	
grievous	36%	24%	30%
cogent	60%	44%	52%
toil	56%	40%	48%
tedious	64%	64%	64%
expeditious	32%	32%	32%
astound	52%	28%	40%
veritably	40%	32%	36%
feeble	48%	32%	40%
afflicted	40%	28%	34%
sluggishly	64%	40%	52%

petrify	32%	20%	26%
lucid	44%	28%	36%
perseverance	28%	12%	20%
folly	52%	36%	44%
rambunctious	28%	24%	26%
mean of GFT scores for list C in each test	45%	32%	

Appendix 6. Main Study Results

T. Graded readers target words spacing and number of participant recognition in the *immediate* LDT AS and the number of the participant production in the *immediate* GFT of the GRG

1.

Graded readers target words spacing and number of participant recognition in the immediate LDT AS and the number of the participant production in the immediate GFT of the GRG.

Consecutive repetition	**LDT AS**	**GFT scores**	**One day spacing**	**LDT AS**	**GFT Score**
1. concur	88%	68%	*1. vicissitude*	72%	52%
2. new-fangled	88%	80%	*2. ingenious*	88%	68%
3. pertinent	76%	48%	*3. proprietorship*	80%	60%
4. detestable	84%	44%	*4. perceptive*	80%	56%
5. misgiving	60%	68%	*5. shrewd*	84%	56%
6. mystification	48%	40%	*6. trespass*	88%	67%
7. obstinate	80%	72%	*7. contrive*	80%	60%
8. colossal	84%	76%	*8. ebullient*	60%	76%
9. unequivocally	80%	60%	*9. idiosyncratic*	80%	64%
10. veritably	84%	40%	*10. acrid*	76%	60%
11. lucid	60%	44%	*11. bonny*	80%	80%
12. sluggishly	88%	64%	*12. apprehensive*	76%	60%
13. astound	50%	52%	*13. brittle*	88%	68%
14. expeditious	64%	32%	*14. feeble*	80%	48%
			15. grievous	72%	36%
			16. cogent	56%	60%
			17. petrify	52%	32%
Two days spacing	**LDT AS**	**GFT score**	**Three days spacing**	**LDT AS**	**GFT scores**
1. contentious	84%	52%	*1. postulation*	80%	52%
2. elucidate	84%	36%	*2. banish*	92%	88%
3. adroit	68%	80%	*3. stupendous*	68%	76%
4. gargantuan	80%	56%	*4. folly*	60%	52%
5. irksome	72%	88%	*5. perseverance*	60%	28%
6. toil	60%	56%	*6. rambunctious*	68%	28%
7. tedious	60%	64%			
8. afflicted	72%	40%			

2.

Timing of occurrence of target words in the graded readers and the number of participant recognition in immediate LDT and number of participant production in the immediate GFT

1st and 2nd	**LDT**	**GFT**	**1st and 3rd**	**LDT**	**GFT**	**1st and 4th**	**LDT**	**GFT**
misgiving	60%	68%	*feeble*	80%	48%	*contentious*	84%	52%
mystification	48%	40%	*vicissitudes*	72%	52%	*adroit*	68%	80%
			ingenious	88%	68%	*gargantuan*	80%	56%
			proprietorship	80%	60%	*irksome*	72%	88%
			ebullient	60%	76%	*toil*	60%	56%
						tedious	60%	64%
1st and 5th	**LDT**	**GFT**	**2nd and 3rd**	**LDT**	**GFT**	**2nd and 4th**	**LDT**	**GFT**
postulation	80%	52%	*concur*	88%	68%	*idiosyncratic*	80%	64%
banish	92%	88%	*new-fangled*	88%	80%	*acrid*	76%	60%
stupendous	68%	76%	*obstinate*	80%	72%	*perceptive*	80%	56%
folly	60%	52%	*colossal*	84%	76%	*shrewd*	84%	56%
perseverance	60%	28%	*veritably*	84%	40%	*trespass*	88%	67%
rambunctious	68%	28%	*lucid*	60%	44%	*grievous*	72%	36%
			sluggishly	88%	64%	*cogent*	56%	60%
2nd and 5th	**LDT**	**GFT**	**3rd and 5th**	**LDT**	**GFT**	**4th and 5th**	**LDT**	**GFT**
elucidate	84%	36%	*contrive*	80%	60%	*unequivocally*	80%	60%
afflicted	72%	40%	*bonny*	80%	80%	*pertinent*	76%	48%
			apprehensive	76%	60%	*detestable*	84%	44%
			brittle	88%	68%			
			petrify	52%	32%			
3rd and 4th	**LDT**	**GFT**						
astound	50%	52%						
expeditious	64%	32%						

Appendix 6. Main Study Results

U. Graded readers target words spacing and the number of participant recognition in the LDT AS and the number of the participant production in the GFT of the GRG in the post-test

1.

Graded reader target words spacing and the number of participant recognition in the LDT AS and the number of the participant production in the GFT of the GRG in the post-test.

Consecutive repetition	**LDT**	**GFT**	**One day spacing**	**LDT**	**GFT**
concur	68%	36%	*vicissitude*	72%	36%
Newf-angled	88%	64%	*ingenious*	88%	44%
pertinent	56%	20%	*proprietorship*	76%	40%
detestable	68%	28%	*perceptive*	72%	32%
misgiving	48%	36%	*shrewd*	72%	56%
mystification	68%	12%	*trespass*	76%	44%
obstinate	64%	28%	*contrive*	60%	40%
colossal	68%	28%	*ebullient*	64%	24%
unequivocally	56%	32%	*idiosyncratic*	60%	16%
veritably	72%	32%	*acrid*	60%	24%
lucid	80%	28%	*bonny*	68%	64%
sluggishly	72%	40%	*apprehensive*	68%	16%
astound	68%	28%	*brittle*	60%	44%
expeditious	72%	32%	*feeble*	80%	32%
			grievous	64%	24%
			cogent	64%	44%
			petrify	80%	20%
Two days spacing	**LDT**	**GFT score**	**Three days spacing**	**LDT**	**GFT**
contentious	60%	32%	*postulation*	68%	12%
elucidate	72%	16%	*banish*	68%	48%
adroit	56%	40%	*stupendous*	60%	52%
gargantuan	76%	28%	*folly*	80%	36%
irksome	64%	44%	*perseverance*	64%	12%
toil	72%	40%	*rambunctious*	64%	24%
tedious	76%	64%			
afflicted	72%	28%			

2.

Timing of occurrence of target words in the graded readers and the number of participant recognition in LDT in the post-test and number of participant production in the GFT in the post-test

1st and 2nd	**LDT**	**GFT**	**1st and 3rd**	**LDT**	**GFT**	**1st and 4th**	**LDT**	**GFT**
misgiving	48%	36%	*feeble*	80%	32%	*contentious*	60%	32%
mystification	68%	12%	*vicissitudes*	72%	36%	*adroit*	56%	40%
			ingenious	88%	44%	*gargantuan*	76%	28%
			proprietorship	76%	40%	*irksome*	64%	44%
			ebullient	64%	24%	*toil*	72%	40%
						tedious	76%	64%

1st and 5th	LDT	GFT	2nd and 3rd	LDT	GFT	2nd and 4th	LDT	GFT
postulation	68%	12%	*concur*	68%	36%	*idiosyncratic*	60%	16%
banish	68%	48%	*new-fangled*	88%	64%	*acrid*	60%	24%
stupendous	60%	52%	*obstinate*	64%	28%	*perceptive*	72%	32%
folly	80%	36%	*colossal*	68%	28%	*shrewd*	72%	56%
perseverance	64%	12%	*veritably*	72%	32%	*trespass*	76%	44%
rambunctious	64%	24%	*lucid*	80%	28%	*grievous*	64%	24%
			sluggishly	72%	40%	*cogent*	64%	44%
2nd and 5th	**LDT**	**GFT**	**3rd and 5th**	**LDT**	**GFT**	**4th and 5th**	**LDT**	**GFT**
elucidate	72%	16%	*contrive*	60%	40%	*unequivocally*	56%	32%
afflicted	72%	28%	*bonny*	68%	64%	*pertinent*	56%	20%
			apprehensive	68%	16%	*detestable*	68%	28%
			brittle	60%	44%			
			petrify	80%	20%			
3rd and 4th	**LDT**	**GFT**						
astound	68%	28%						
expeditious	72%	32%						

Appendix 6. Main Study Results
V. The analysis of the three groups

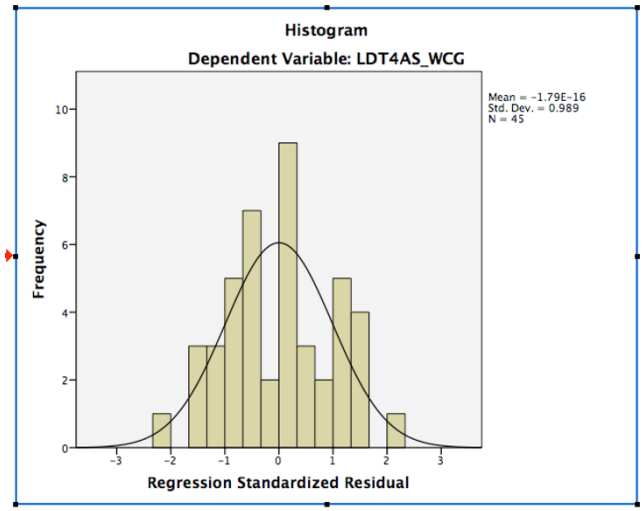


Figure 1 Histogram of standardized residuals of multiple regression analysis of the LDT AS scores of the RMCG and the frequency of exposure to the target word

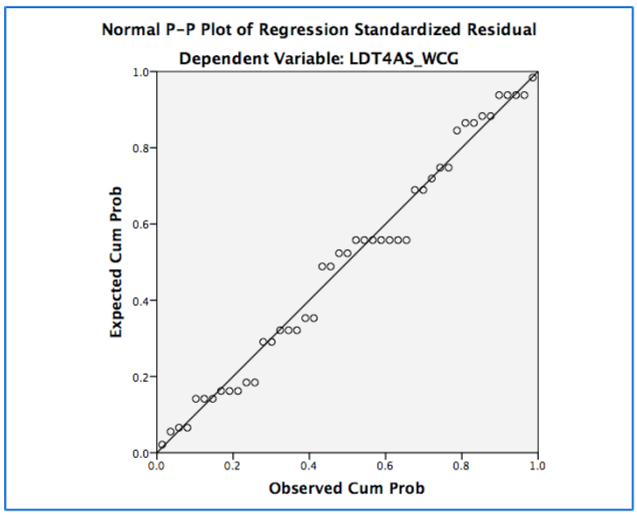


Figure 2 P-P plot of standardized residuals of multiple regression analysis of the LDT AS of the RMCG and the frequency of exposure to the target word

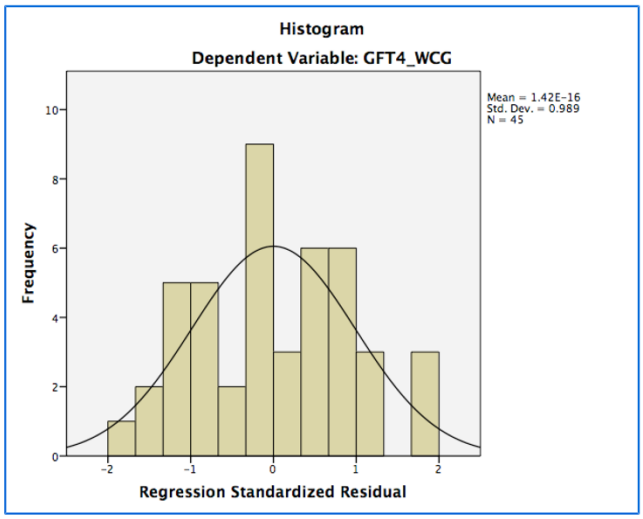


Figure 4 Histogram of standardized residuals of multiple regression analysis of the GFT scores of the RMCG and the frequency of exposure to the target word

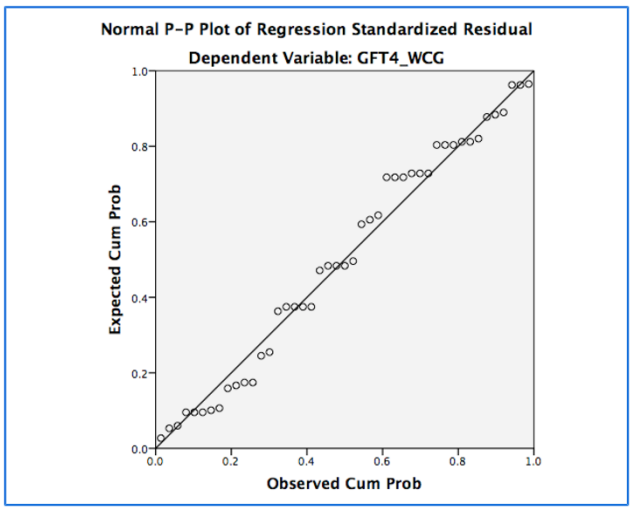


Figure 4 P-P plot of standardized residuals of multiple regression analysis of the GFT scores of the RMCG and the number of frequency of exposure to the target word

5.

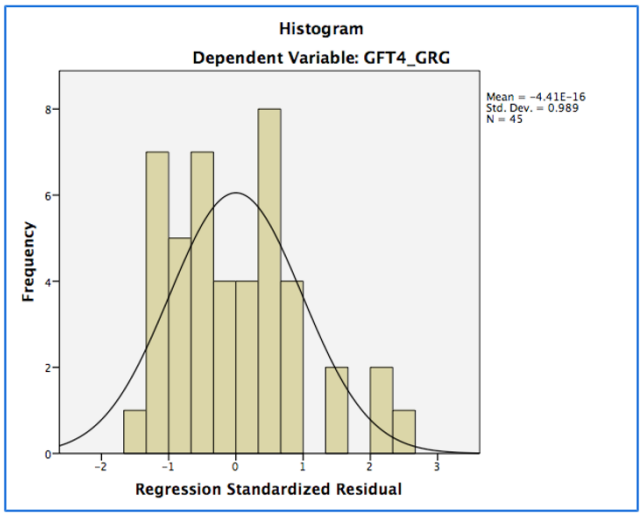


Figure 5 Histogram of standardized residuals of multiple regression analysis of the GFT scores of the GRG and the number of syllables of the target word

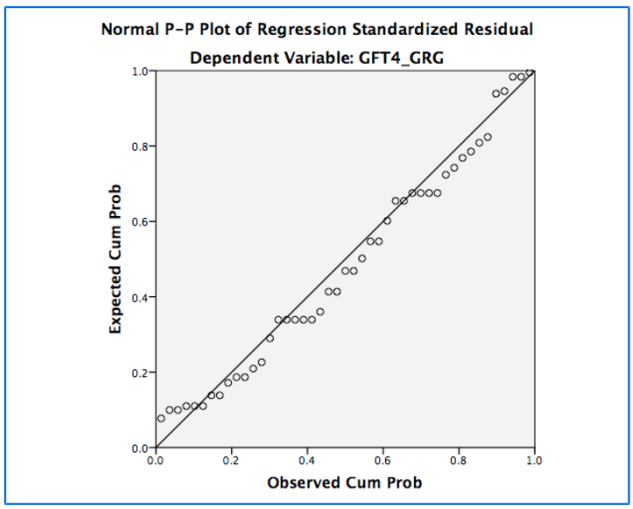


Figure 6 P-P plot of standardized residuals of multiple regression analysis of the GFT scores of the GRG and the number of syllables of the target word

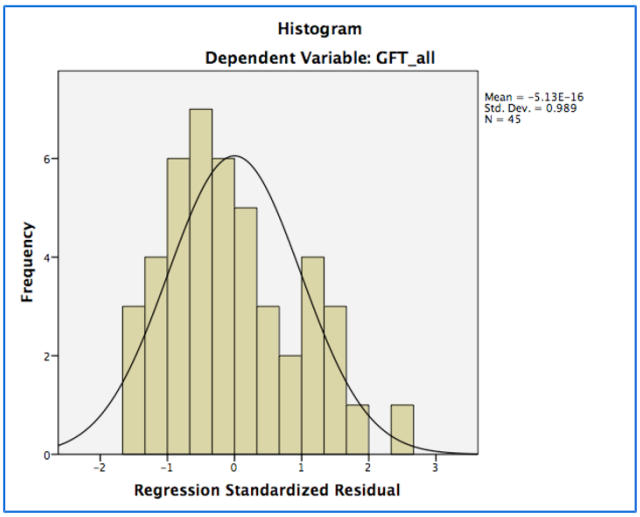


Figure 7 Histogram of standardized residuals of multiple regression analysis of the GFT scores of all groups and the number of syllables of the target word

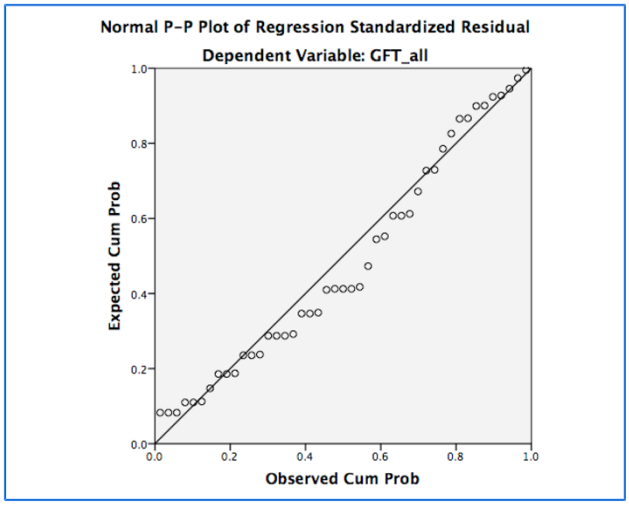


Figure 8 P-P plot of standardized residuals of multiple regression analysis of the GFT scores of all groups and the number of syllables of the target word