



**EXPLORING THE EFFECTIVENESS OF EXTENSIVE
READING ON INCIDENTAL VOCABULARY ACQUISITION
BY EFL LEARNERS: AN EXPERIMENTAL CASE STUDY IN
A LIBYAN UNIVERSITY**

By

Hamed Alahirsh

Thesis submitted to the University of Nottingham for the degree of Doctor
of Philosophy

July, 2014

Acknowledgements

First and foremost, thanks should be expressed to my God (Allah) for his assistance and for giving me the opportunity of being engaged in advanced studies in the UK. I also should thank him for letting me have the courage to overcome all the difficulties I encountered throughout the period of my doctoral research.

I would like to express my most sincere thanks to my supervisor Dr Barbara Sinclair for her precious supervision, inspiring guidance, constant encouragement, and for being patient in solving all the research obstacles and difficulties I encountered during this very long journey. Special thanks must also go to my co-supervisor Dr Juliet Thondhlana for her valuable supervision and constructive feedback on my thesis chapters. I have been very fortunate to experience the inspiration of excellent teachers and supervisors whom I have learnt from them many key skills related to my research and writing this thesis. Unfortunately, Dr Pemberton who started this journey with me is no longer with us to celebrate the completion of my thesis.

I am especially indebted to Jane Restorick for her appreciated assistance in using the SPSS software and carrying out all the statistical work required for the data analysis of this study. Moreover, I am grateful to many people in the research school. First, I want to give my great thanks to the Head of the English Language Department for his overwhelming support of and assistance with this study. He made all student databases easily available for the purpose of sampling in my research. As for data collection, my deepest appreciation must go to the department teachers, who offered valuable assistance in administrating and marking the vocabulary tests. Thanks should also be expressed to the two university librarians who administrated the extensive reading (ER) library throughout the two-month programme. Furthermore, I am deeply grateful to the department students who participated in both the piloting procedures and this experimental research. I thank them for their patience and co-operation. Without them, this study would not have been successful.

Finally, my deepest appreciation must go to my wife and my children without whose support I would never have succeeded. They shared the highs and lows of this seemingly epic journey with me and hardly ever complained during my absence. I should also thank some people back in Libya including my brothers, sisters, relatives, and friends for their constant emotional support and encouragement.

Dedication

To my late father and mother

*To the soul of Dr. Richard Pemberton, who started this journey as my
main supervisor, but died halfway.*

Abbreviations

The abbreviations listed below are used in this thesis

- EFL: English as a foreign language
- ESL: English as a second language
- ELT: English language teaching
- EAP: English for academic purposes
- L1: First language
- L2: Second language
- ER: Extensive reading
- P1 to P18: Participants in the experimental research group
- C1 to C18: Participants in the control group
- VLT: The Vocabulary Level Test
- 1K: The first 1,000 most frequently used words in English
- 2K: The second 1,000 most frequently used words in English
- Off-list: Words beyond the second 1,000 words
- AWL: Academic Word List
- IH: Input Hypothesis

Abstract

Considerable worldwide research has investigated incidental vocabulary learning from L2 reading, yet so far nothing has been published about the actual learning that comes from reading various texts. This study investigated incidental lexical growth and retention by Libyan university EFL majors who were involved in a two-month ER programme. Their vocabulary gain was measured 1 week after the participants completed the Extensive Reading (ER) programme, 2 weeks later and 9 months later. The value of this study is that it used an innovative approach, which was developed from a research design by Horst (2005). This included the electronic scanning of books and lexical frequency profiling, helping the researcher to create individualised corpus profiles from the entire set of different texts the participants read. This data was then used to select target words for each participant.

The methodology was an experimental case study, which entailed an experimental and control group design. The participants were EFL learners who studied English as their subject of specialisation in one of eight Libyan state universities. An original number of 80 participants were randomly selected from the entire population in the English Language Department and assigned equally between the experimental and control groups. However, due to the fact that this study was carried out in very anomalous circumstances (during the Libyan uprising, which started on 15/02/2011), the number of participants who successfully completed the ER programme was affected (18 participants in each group).

The study showed that by using a corpus analysis strategy, it was feasible to measure learners' individualised pre-post treatment acquisition of the vocabulary they encountered in a large number of ER graded texts. The findings of the study demonstrate that ER significantly improved the Libyan EFL learners' incidental vocabulary acquisition. By the end of the study, it was found that about the third of the target words had been acquired by the participants at both receptive and productive levels of knowledge. The findings

further indicate that word repetition was an important factor for an incidental pick up of vocabulary from the ER. In relation to long-term retention rate of learning, the findings suggest that incidental word knowledge, acquired through ER, significantly declines over time.

Table of Contents

Acknowledgements	I
Dedication	III
Abbreviations	IV
Abstract	V
Chapter 1: Introduction	1
1.1 Introduction	1
1.2 Background to the study.....	1
1.3 Statement of the problem	5
1.4 The context of the study	8
1.4.1 The geographical and sociolinguistic background.....	9
1.4.2 Education system in Libya.....	9
1.4.2.1 Early stages of education	9
1.4.2.2 Current system of Libyan education in Libya.....	12
1.4.3 Background of ELT in Libya	18
1.4.4 Current status and challenges of ELT in the Libyan universities	22
1.5 The study: rationale, aims, and research questions	24
1.6 Significance of the study	26
1.7 Structure of the thesis	27
Chapter 2: Review of literature	30
2.1 Introduction	30
2.2 L2 vocabulary acquisition	30
2.2.1 What constitutes word knowledge?	32
2.2.2 Receptive and productive word knowledge	35
2.2.3 Incidental vs. intentional vocabulary learning	37
2.3 L2 Reading theories.....	44
2.3.1 Defining reading	45
2.3.2 Models of L2 reading processes.....	46
2.3.2.1 Bottom-up processing in reading	46
2.3.2.2 Top-down processing in reading.....	49
2.3.2.3 Interactive processing in reading	52
2.3.3 Problems faced by Arab EFL learners when reading in English	56
2.3.3.1 Problems of misunderstanding of the reading process.....	57
2.3.3.2 Problems of insufficient competence in and use of English	59
2.3.3.3 Reading problems due to the differences between Arabic and English	60
2.3.3.4 Problems due to the English spelling/ sound system	61
2.3.4 Extensive reading.....	63
2.3.4.1 Definitions.....	63
2.3.4.2 Theoretical background.....	65
2.3.4.3 Characteristics of extensive reading.....	66
2.3.4.4 Some issues with setting up an ER programme	72
2.3.4.5 Graded materials for extensive reading.....	73

2.3.4.6	Benefits of extensive reading	74
2.3.4.7	Drawbacks of extensive reading	77
Chapter 3:	Incidental vocabulary research	80
3.1	Introduction	80
3.2	Studies on L1 (First Language) acquisition	80
3.3	Studies on L2 (Second Language) acquisition	83
3.3.1	Incidental vocabulary learning from reading short texts.....	83
3.3.2	Incidental vocabulary learning from ER.....	98
Chapter 4:	Methodological considerations and pilot study	116
4.1	Introduction	116
4.2	Ontological and epistemological perspectives	116
4.3	Research methodology	118
4.4	Selecting and designing paradigms	119
4.5	Objectives of the study	122
4.6	Research hypotheses.....	122
4.7	Research questions	123
4.8	Ethical considerations.....	123
4.9	Pilot study.....	127
4.9.1	Pilot subjects	128
4.9.2	Piloting stages and results	128
4.9.2.1	Determining levels of reading materials	128
4.9.2.2	Determining target 2K words.....	130
4.9.2.3	Piloting reading programme and vocabulary testing procedures	134
Chapter 5:	Research design (main study).....	140
5.1	Introduction	140
5.2	The experimental design in detail	140
5.2.1	Research context	142
5.2.2	Participant description.....	143
5.2.3	Participant selection	144
5.2.4	Tasks for the two groups.....	150
5.3	Research procedures	151
5.3.1	Participants' choices of books.....	151
5.3.2	Scanning (corpus work)	153
5.3.3	Target words	158
5.3.4	Library and readers	162
5.3.5	The treatment	164
5.4	Research instruments for data collection.....	169
5.4.1	The vocabulary test	169
5.4.1.1	Test administration.....	169
5.4.2	Semi-structured interviews	177
Chapter 6:	Overall results.....	182
6.1	Introduction	182
6.2	Data presentation, analysis, and interpretation.....	182

6.3	Research Question 1: <i>Does reading extensively enhance adult, Libyan, EFL learners' vocabulary knowledge?</i>	183
6.3.1	Experimental vs. control groups' prior knowledge of target words.....	183
6.3.1.1	Overall pre-existing target word knowledge.....	186
6.3.1.2	Pre-existing off-list word knowledge.....	187
6.3.1.3	Pre-existing 2K word knowledge.....	187
6.3.2	Experimental vs. control groups' target word knowledge improvement... ..	190
6.3.2.1	Overall incidental vocabulary gain	190
6.3.2.2	Off-list words gain	200
6.3.2.3	2K words gain	204
6.3.3	Summary of main findings.....	208
6.4	Research Question 2: <i>Does word repetition affect receptive and productive incidental vocabulary learning from extensive reading? If so, how many exposures are needed for incidental vocabulary learning?</i>	216
6.4.1	Receptive knowledge improvements	217
6.4.2	Productive knowledge improvements	220
6.4.3	Summary of main findings.....	222
6.5	Research Question 3: <i>How much of the newly learnt vocabulary is retained 9 months after the ER programme?</i>	227
6.5.1	Comparison between learning retention scores over nine months.....	227
6.5.2	Absolute gain retention over nine months	233
6.5.3	Summary of main findings.....	236
Chapter 7:	Individual performance data.....	245
7.1	Introduction	245
7.2	Research Question 4: <i>Will individualisation in the selection of ER texts reveal any significant difference in the number of vocabulary items learned as well as in the learners' perception of the ER programme?</i>	245
7.2.1	Individual data analysis of experimental vs. control participants' learning.....	246
7.2.2	Incidental vocabulary growth from individualised reading	251
7.2.3	The participants' perceptions of the ER programme	266
7.2.4	Summary of main findings.....	269
Chapter 8:	Conclusions, contributions, limitations, and implications	279
8.1	Introduction	279
8.2	Contributions to knowledge 1: Major Findings	279
8.2.1	Vocabulary development	280
8.2.2	Receptive and productive growth	280
8.2.3	The relationship between frequency of word encounters and learning.....	281
8.2.4	Attrition rates in incidentally learnt vocabulary	282
8.2.5	Individualised reading and incidental vocabulary growth	282
8.2.6	Learners' attitudes and motivation.....	283
8.3	Contributions to knowledge 2: Innovation in research design	283
8.3.1	Individualised assessment	283
8.3.2	Delayed measurements of incidental vocabulary learning.....	285
8.3.3	New research in the Libyan EFL context.....	286

8.4	Limitations of the study.....	286
8.5	Suggestions for further research.....	289
8.6	Pedagogical implications.....	290
8.7	Empirical implications	295
8.8	Personal reflections	296
References		299
Appendices		322
Appendix 1:	Oxford bookworm levels tests (stages 3, 4,and 5)	323
Appendix 2:	Instant book report.....	326
Appendix 3:	Schmitt’s (2000) Vocabulary Levels Test.....	328
Appendix 4:	Experimental vs. control participants’ scores on VLT.....	337
Appendix 5:	List of the target Off-list words.....	338
Appendix 6:	List of the target 2K words.....	339
Appendix 7:	Sample of reading materials	340
Appendix 8:	Time record sheet	343
Appendix 9:	Non-parametric Wilcoxon Signed Ranks Test.....	344
Appendix 10:	Sample of book glossary (<i>The Railway Children</i>)	355

List of Tables

Table 1-1: Enrolment in university education after gaining the specialisation secondary education certificate	15
Table 2-1: Aspects involved in word knowledge	34
Table 4-1: Bookworms Graded Scheme (Oxford University Press).....	129
Table 4-2: Sample of 2K word families (Stage5 books).....	131
Table 4-3: The number of checklist tests at each stage.....	132
Table 4-4: Self-reported checklist test Sample 2K items.....	133
Table 4-5: Results of pilot 2K words.....	134
Table 4-6: Individualised test sheet	135
Table 4-7: Piloting the time required for reading a graded book.....	137
Table 5-1: Types of experiments (Nunan, 1992; p. 41).....	141
Table 5-2: The number of four academic year students in the English department ..	145
Table 5-3: Output of Vocabprofile analysis of entire words in the graded books	157
Table 5-4: The target words and the number of their use across the individualised vocabulary tests.....	162
Table 5-5: The number of times each of the experimental books circulated around the participants	168
Table 5-6: Guiding questions and prompts used in participants' interviews	179
Table 6-1: Test of Normality	185
Table 6-2: Comparison of the experimental and control groups for the overall pre- vocabulary test scores assessed by means of the Paired-Samples t-Test	186
Table 6-3: Comparison of the experimental and control groups on pre-test off-list item scores, assessed by means of the Paired-Samples t-test.....	187
Table 6-4: Comparison of the experimental and control groups on pre-test 2K items scores, assessed by means of the Paired-Samples t-test.....	187
Table 6-5: Percentages of words rated at the each level of the scale in the pre-test by the experimental and the control groups	188
Table 6-6: Tests of Normality.....	191
Table 6-7: Comparison of the experimental and control group post-vocabulary test scores, assessed by means of the Paired Samples t-test	191
Table 6-8: Mean comparison of the experimental and control groups on the pre- and post-tests.....	192
Table 6-9: Mean percentage difference for words rated by the experimental group at each the four word knowledge scale levels in the pre-and post-test.	193
Table 6-10: Mean percentage difference of words rated by the control group at each of the four word knowledge scale levels in the pre-and post-test.....	195
Table 6-11: The experimental and control group's target word knowledge improvement	197
Table 6-12: Comparison of the experimental and control group for post-test off-list item scores, assessed by means of the Paired-Samples t-test.....	200
Table 6-13: Comparison of the experimental and control group levels of knowledge of off-list items in the pre-and post-test	201
Table 6-14: The experimental and control group's knowledge improvement for off-list words.....	203

Table 6-15: Comparison of the experimental and control group on post-test 2K items scores assessed by means of the Paired-Samples t-Test	204
Table 6-16: Comparison of the experimental and control group levels of knowledge 2K items in the pre-and post-tests.....	205
Table 6-17: The experimental and control group's gain of 2K words	207
Table 6-18 : Receptive gain percentages according to each frequency groups	217
Table 6-19: Spearman's correlation between the number of target word repetitions and incidental receptive learning	219
Table 6-20: Productive gain percentages according to each frequency group.....	220
Table 6-21: Spearman's correlation between target word repetitions and productive incidental learning.....	221
Table 6-22: Descriptive statistics for the vocabulary test scores over time	228
Table 6-23: The mean word percentages by level of knowledge over time	230
Table 6-24: Changes in target word knowledge nine months after the first post-test	234
Table 6-25: Mean score differences for off-list and 2K words in the post and delayed test, shown using a Paired-Samples t-test.	236
Table 6-26: Word frequency and incidental learning retention rates after nine months	240
Table 7-1: Individualised learning comparison between participants' post-test gains for the experimental group and their corresponding subjects in the control group	248
Table 7-2: Levels and numbers of participants and their selected books, numbers of ER weeks and amount of reading carried out per week.....	253
Table 7-3: The amount of participants' extensive reading and number of words included in their individualised tests.....	254
Table 7-4: Experimental participants and their individualised reading and incidental vocabulary acquisition rates.....	256
Table 7-5: Spearman's correlation between the percentage of actual read texts, reading per week, and the percentage of incidental vocabulary gain.....	257
Table 7-6: Comprehension rates of different graded stories read by the experimental participants	261
Table 7-7: Spearman's correlation between the participants' incidental vocabulary learning percentages and the level of texts comprehension	262
Table 7-8: Learners' reported vocabulary learning strategies for discovering the meaning of the new words they met in the graded books	264
Table 7-9: Spearman's correlation between vocabulary gain and strategies used to determine word meaning.....	265

List of Figures

Figure 1-1: The order of stages in the Libyan educational system (adapted from Asker, 2011)	13
Figure 4-1: Types of research design in the language learning field (from Grotjahn, 1987)	120
Figure 5-1: An example of the process of converting image (pdf) to text (document)	154
Figure 5-2: Samples of Vocabprofile output of word frequency categorisation (The riddle of sands).....	155
Figure 5-3: VocabProfile analysis (overall output) of the entire text of The Big Sleep	156
Figure 6-1: Percentage of incidental receptive and productive learning by frequency category	222
Figure 6-2: Mean score over time	228
Figure 6-3: Mean percentage of word levels of knowledge over time	231
Figure 6-4: Mean percentage of participants' receptive and productive knowledge over time	233
Figure 7-1: Percentages of target words gained for each pair after the treatment	247
Figure 7-2: The experimental groups' post treatment individualised active knowledge scores.....	251

Chapter 1:

Introduction

Chapter 1: Introduction

1.1 Introduction

This introductory chapter provides an overview of the current study. It begins with some brief background information about the study and then states the research problem, before describing the context in which it was conducted. The chapter proceeds by mentioning the rationale and aims for undertaking the research and listing the research questions which were addressed in this study. Next, the significance of the study is presented, and the structure of this thesis is also outlined. On the whole, this chapter aims at giving the readers a holistic picture, before the research theme is elaborated on in the subsequent chapters.

1.2 Background to the study

There is a widespread recognition that vocabulary is a central and key element of successful second language (L2) learning. However, one of the major difficulties that L2 learners of English encounter, lies in the enormous number of words they have to learn (Laufer, 1992). A considerable number of studies have investigated the number of words English language learners need to know in order to be able to communicate effectively in their target language. To obtain some indication of the size of the vocabulary task facing L2 language learners, researchers have been primarily concerned with discovery of size of a native speaker's vocabulary. Findings have estimated that a five-year-old native speaker knows around 4000 to 5000 word families¹, and by the time s/he graduates from the university, s/he will have acquired around 20,000 word families (Goulden et al., 1990).

¹ Word family is a base word with its inflections and derivatives (stimulate + stimulated, stimulates, stimulating, stimulation, and stimulative) (Schmitt & McCarthy, 1997, p. 331)

Taking these figures into account, Nation (2006) concludes that English as a foreign/second language (EFL/ESL) learners need to know as many as 8,000-9,000 word families to be able to comprehend 98% of a variety of English texts, and a slightly smaller number of about 6,000 to 7,000 word families to gain the same percentage of comprehension from spoken input. Laufer (1992) suggests that in order to attain this threshold level of unmodified text comprehension, L2 learners need to know a minimum of 3,000 of the most frequently used word families. However, other researchers suggest a higher estimation of 5,000 base vocabulary items (Nation, 1990; Laufer, 1997). This number of words is believed to be sufficient to meet the 98% vocabulary coverage recommended by Hu and Nation (2000) for unassisted comprehension of a specialised text.

Research findings from L1 speakers indicate that since explicit vocabulary learning cannot account for a significant amount of the vocabulary which learners acquire, the majority of L1 vocabulary must be acquired incidentally as a result of frequent exposure. This assumption can also be applied to L2 learning. In fact, it is generally acknowledged that language learners may not be able to acquire a substantial enough level of vocabulary knowledge solely through explicit instruction, due to the fairly short periods of classroom teaching. This point is particularly crucial, since the vocabulary teaching methods used in most EFL classrooms are, in general, “not very effective and even demotivating” (Schouten-van Parreren, 1989, p. 75). Thus, it has been pedagogically recommended that classroom instruction should comprise the most frequent words in English and all words beyond these frequency bands should instead be encountered by learners developing their own strategies to learn the low frequent word families (Nation, 1990; Coady, 1993). In this respect, reading has been considered an effective means for promoting incidental growth of L2 vocabulary. One major justification for such an

assumption is that written discourse tends to contain higher type-token ratios² than spoken discourse, which usually comprises more frequent vocabulary and lower type-token ratios (Coady, 1997; Schmitt, 2000). Therefore, “it would be optimistic to expect to learn a wide range of vocabulary from only spoken discourse” (Schmitt, 2000, p. 150).

A growing body of literature acknowledges the significance of reading as a means of vocabulary development in EFL learners. As a result, a number of studies have been conducted to investigate incidental vocabulary acquisition through reading in both ESL and EFL contexts. Findings from many of these studies suggest that a substantial amount of L2 lexis can be incidentally gained through repeated encounters with the target vocabulary (Krashen, 1989b; Day et al., 1991; Ellis & He, 1999; Gass, 1999; Huckin & Coady, 1999; Rott, 1999b; Waring & Takaki, 2003; Horst, 2005; Brown et al., 2008). In fact, a number of these studies have reported significant improvement in the subjects’ vocabulary knowledge after they were given extensive reading (ER) materials (Dupuy & Krashen, 1993; Horst, 2005; Pigada & Schmitt, 2006; Al-Homoud & Schmitt, 2009). However, despite being widely cited in the literature, findings gained from many of these large-scale studies have been criticised as having serious methodological limitations. A detailed discussion of the methodological problems of previous incidental vocabulary acquisition research will be presented later in section 3.3.

According to Krashen’s Input Hypothesis (IH) (1985, 1991), input should be comprehensible in order for language acquisition to take place. This implies that learners need to know a large amount of vocabulary in the language they are learning in order to comprehend non-mediated second language texts. That is to

² Type-token ratios is a measure of vocabulary variation within a written text or a person’s speech. A high Type-token ratio indicates a large amount of lexical variation and a low type-token ratio indicates relatively little lexical variation.

say, lexical acquisition is central to second language acquisition (Coady & Huckin, 1997); therefore, if a learner has inadequate vocabulary, s/he will not be able to comprehend the language they are encountering. With regard to this notion, Cook (1996, p. XI) emphasises that vocabulary knowledge is a basic requirement of learning communication, and that “no amount of grammatical or other type of linguistic knowledge can be employed in communication or discourse without the mediation of vocabulary”. In a similar vein, Read (2000, p. 14) states that “vocabulary is not just a set of linguistic units but also an attribute of individual language learners, in the form of vocabulary knowledge and the ability to access that knowledge for communicative purposes”. This suggests that increasing vocabulary knowledge should be a basic part of the process of language learning. Therefore, learners need to develop their lexical knowledge through instruction or, more effectively, perhaps, through incidental word learning from ER.

With reference to the importance of reading as a means of increasing knowledge of L2 vocabulary, Nation (2001, p. 155) emphasises that “the use of reading and other input sources may be the only practical options for out of class language development for some learners”. As is succinctly pointed out by Nation and Wang (1999), in order to effectively achieve its pedagogical goals, ER should involve L2 learners in reading graded readers because such materials provide input appropriate to the learners’ level of proficiency. This is consistent with the assumption of IH, that we acquire both spelling and vocabulary when we are exposed to comprehensible input (Krashen, 2004).

As in many similar EFL contexts, the teaching and learning of vocabulary through ER has previously been undervalued in the Libyan classroom. However, with the increasing attention that ER has received worldwide as an effective approach for developing different components of L2 language (Day & Bamford, 1998), it has become important to introduce this approach in Libya and to investigate its effects on Libyan EFL learners.

1.3 Statement of the problem

English in Libya is taught as a foreign language, so the classroom is usually the only source of input for EFL Learners. Classroom instruction is usually restricted to the syllabus which is provided by the department and, as a result, students often appear to suffer the handicap of failing to find supplementary materials to improve their English language skills. From my experience as a teacher in a Libyan university, I would also strongly argue that, in addition to the teaching methods adopted by these EFL teachers, factors such as the teaching materials, the learners' levels of proficiency, the learners' use of learning strategies and their level of motivation may play a key role in determining their success in improving their EFL language skills. Because the present study is investigating incidental vocabulary acquisition from reading in Libya, the discussion in this section will be limited to two aspects of language education: vocabulary learning and reading.

As has already been argued, vocabulary acquisition has consistently been regarded as one of the most important components in learning a second language (Meara, 1980; Schmitt, 2000; Nation, 2001). This important concept is stressed by Lewis (1993, p. 89) who sees vocabulary as "the core or heart of language". Schmitt (2010) further argues that the importance of vocabulary in language learning is recognised by learners, who "carry around dictionaries and not grammar books" (p.4). Nevertheless, this important language skill has not yet gained a status of priority among the curricula of English departments in Libyan universities. Working as an EFL teacher in the English department of such a University, I noted that the problems experienced by students in learning English seemed to be related principally to their lack of sufficient knowledge of English vocabulary. Students very often struggled not only to communicate in English, but also to read and understand some of the materials provided by their teachers. Although classroom activities are currently designed to teach grammatical rules and to develop the learners' range of EFL vocabulary through memorisation, I have observed that in Libyan Universities, learners with English

as a major frequently struggle to recall the meanings of even words that are in common usage. Therefore, the more traditional methods of vocabulary instruction, which focus on memorisation of isolated words, seem not to have been effective (Orafi & Borg, 2009). In addition, Nation (1990, p. 19) stresses that “the learning burden of a word can be affected by the way it is taught [...] as a result of bad organisation the difficulty of learning a word is increased”. The results of a previous investigation (MA dissertation) of vocabulary acquisition, carried out by the present researcher (Alahirsh, 2006), indicated that both Libyan EFL teachers and learners lacked knowledge of the different approaches for improving learners’ lexical knowledge.

Despite the growing global trend for EFL instructors and researchers to use an ER approach to teach different language skills (Janopoulos, 1986; Robb & Susser, 1989; Hafiz & Tudor, 1990; Lai, 1993; Cho & Krashen, 1994; Masuhara et al., 1996; Tsang, 1996; Mason & Krashen, 1997; Kusanagi, 2004; Taguchi et al., 2004; Al-Homoud & Schmitt, 2009), the intensive approach still plays the dominant role in EFL reading classes in Libya. On these courses, the syllabus and timetables are decided by the Head of the English department, whilst the teacher selects the texts or the textbooks for their students. Class reading activities are mainly designed and planned to improve the comprehension skills of the students, rather than to improve their vocabulary. In my experience, the two most common classroom reading approaches in Libyan university English departments are either the teacher reading a passage aloud to the whole class, or the teacher asking the students to read aloud sections of a passage in turn. In both cases, there is minimal reading practice in English. In this case, the role of the teacher is to check the students’ pronunciation, understanding of the general meaning of the passage, and then to ask comprehension questions which require short answers. The teacher also usually asks the students if they have met any unknown words in the passage and then tries to translate them into Arabic. In my experience as an EFL teacher in the aforementioned context, I noticed that many students felt uncomfortable when speaking in front of their classmates,

and therefore sometimes showed a reluctance to participate in the reading activities set.

In such a highly prescriptive context, learners may be confronted with a number of barriers to understanding the texts and enjoying their reading. Furthermore, Libyan universities do not allow the teachers to set any additional materials for their students. Instead teachers have to adhere to the syllabus and texts provided by the university, despite the fact that they are sometimes aware of the inappropriateness of the textbooks chosen for their students.

With regard to this situation, several issues which have emerged from these reading classrooms appear to be problematic. In my experience as one of those who has been involved in teaching reading comprehension to undergraduate students in an English Language Department in Libya, it seems that reading materials are not always enjoyed by students, and there are many occasions when students feel bored in their reading classes. As a result, traditional approaches to teaching reading in Libya do not seem to be beneficial. Furthermore, it is common practice for teachers to choose the textbooks they believe to be easy for their students to understand, so as to avoid demotivating their learners with harder texts. However, little attention is given to the students' interest in the reading topics. In addition, most reading classes involve students in reading one long story (for example, *Oliver Twist*, by Charles Dickens) throughout a whole academic year, although, for many of them, reading about various topics might be more interesting and motivating.

Unfortunately, the students in these English departments have to adhere to the teachers' materials for examination purposes, yet generally, they seem to struggle to obtain great benefits from such reading activities, especially in terms of vocabulary improvement (Al-Naiely, 2012). Even if the students intend to find alternative texts to practise individual reading with, they often encounter a severe lack of EFL materials. In reference to his experience as a foreign EFL

teacher in a Libyan university, Rajendran (2010, p. 63) discusses this problem, stating that he found:

[a] lack of interest and motivation and competitive spirit on the part of students, non-availability of study materials, books, journals and newspapers, absence of scholarly ambience and some of the blocks which hinder the process of learning the English language.

This point has also been noted by Najeeb (2012), who concludes that few Libyan students appear to have the habit of reading - a fact that is accentuated by the absence of appropriate English materials.

To sum up, the challenges faced by both EFL teachers and learners in learning and teaching English vocabulary and reading suggest that urgent steps need to be taken to address these challenges in Libyan University EFL departments. The worldwide interest surrounding research into vocabulary learning has stemmed from a recognition of the importance of exploring the most effective ways of developing L2 learners' vocabulary knowledge. Hunt and Beglar (1998) suggest three approaches to vocabulary learning: incidental learning, explicit instruction and independent strategy development. Among these three approaches, incidental vocabulary learning is considered the most important for L2 learning (Hong, 2010). This incidental learning can follow verbal input through spoken communication or by reading. The latter is seen as an efficient approach for L2 teaching because these two language learning activities (reading and incidental vocabulary acquisition) can occur concurrently (Huckin & Coady, 1999).

1.4 The context of the study

This section provides an overview of the context in which this study took place, including a general historical background of the education system in Libya and some information about English language teaching and learning in Libya, with particular focus on ELT at university level.

1.4.1 The geographical and sociolinguistic background

Libya is an Arab country belonging to the African continent and is bordered to the north by about 2000 kilometres of coast along the Mediterranean Sea. As reported by Bugaighis (2011), the country has a total area of 1.8 million square kilometres, which makes it the fourth largest country by area in Africa (Tantani, 2012). It shares its western border with Egypt, its eastern border with Tunisia and Algeria, and its southern borders with Sudan, Niger, and Chad (Bugaighis, 2011). The total population consists of about 6 million inhabitants (LNTTO, 2012), 90% of whom are ethnically Arab. As a result of this, it is considered a monolingual country, regardless of the fact that 10% of the inhabitants speak Berber as their L1 and Arabic as their L2 (Asker, 2011). Apart from English, no second foreign languages are spoken by the Libyans. In fact, although the country was occupied by Italy from 1911 to 1942, this does not appear to have had a noticeable impact on the Libyan culture or the present sociolinguistic structure. Generally, very few Libyan people, most of whom are elderly, speak fluent Italian (Asker, 2011).

1.4.2 Education system in Libya

This study is concerned with Libyan students who study English as their main subject at university level. Before giving some background information about EFL teaching and learning in Libya, it is worth providing, first, a brief historical background of the Libyan education system and discussing the stages of education leading up to university.

1.4.2.1 Early stages of education

Whilst reviewing the education system in Libya, it is worth including a discussion of its foundation. The establishment of the Libyan education system is rooted in the era of Islamic conquest, which began in 642 AD (Obeidi, 2001). In her review of the establishment of the education system in Libya, Obeidi (2001) reported that during these very early stages of history, education was

principally voluntary and was based on religious teaching and Arabic literacy classes. The system of education entailed two main levels, both of which were largely attended by boys. Girls had very little opportunity to gain an education, due to the family and cultural constraints that dominated the society at that time. At the first level, 6-7 year old children could attend a Quran school called a 'Katatib'. The 'Katatib' consisted of classes of children who took lessons in how to read and write the Arabic script in classrooms and which were typically affiliated with mosques. The second level of education was more advanced than the 'Katatib', and male children at this stage had to memorise the whole of the holy Quran and learn the 'Sharieaa' (Islamic rules) under the supervision of specialist scholars. Nevertheless, this type of education was provided by a very limited number of mosques, which were mainly situated in only a few cities. Upon completion of this level, students would be legally qualified to participate in the dissemination of primary level religious education across the country (Obeidi, 2001).

Between 1551 and 1911, Libya was under the control of the Othman Empire. During this period the 'Katatib' continued to play a key role in education, although the Othman era also witnessed the introduction of state schools. What may have caused the 'Katatib' to maintain their role, is the fact that the Othman authorities established only a very small number of state schools, mainly in Tripoli. Like the Katatib, these state schools offered lessons to males and females in basic literacy, Arabic language, the Quran, and other religious principles (Obeidi, 2001; Evered, 2012). In addition to the establishment of state schools, a higher level training system was founded by Othman authorities, with a limited number of institutes, such as the Murad Pasha and the Darghut Pasha, which mainly provided courses in Islamic law. At the same level of education, 'Zawaya' (Islamic centres) also offered teaching courses in history, mathematics, geography, and medicine (Obeidi, 2001).

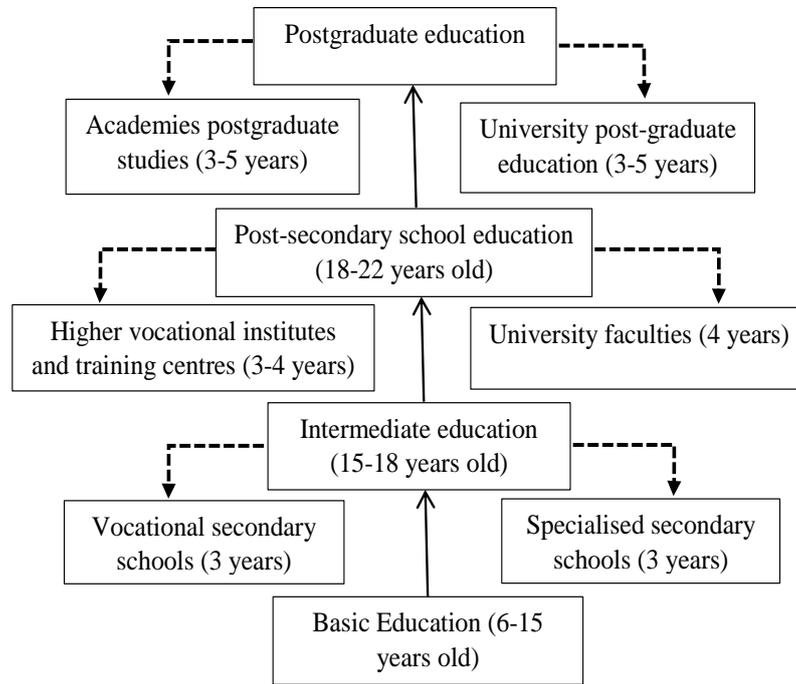
In 1911, Italy invaded Othman Libya and occupied the country for about 3 decades (1911-1942). According to the Libyan National Report (GPCE, n. d), when Italian invaders landed in Libya, they closed down all existing Othman schools and Quranic learning centres and burnt down the remaining Islamic schools and libraries. Gulinao (2013) wrote that by taking this step, the Italian authority aimed at re-establishing a new education system which would coincide with the Italian invasion policy in Libya. Thus, in 1939, about 100 schools were constructed across the country. However, ordinary Libyans in both urban and rural areas were not allowed to attend these schools (Gulinao, 2013). In fact, as is stated in the Libyan National Report (GPCE, n. d), only 5% of Libyan people were permitted to join these Italian schools. The classroom instruction in these schools was provided through the Italian language, which was a compulsory subject in schools, with Arabic being considered a second language in the country. Libyan pupils studying in these schools were not allowed to study beyond the 4th year at the primary level. As a consequence of such a policy, the majority of the country was left illiterate, as few children were able to obtain a basic education (GPCE, n. d).

According to Obeidi (2001), education in Libya made a certain level of recovery when the allied French-British occupation took over from the Italians in 1943, ending in 1951. At this time Libya was divided into three regions: Tripoli and Cyrenaica in north and Fezzan in the south. While Tripolitania and Cyrenaica remained under British administration, the French ruled in the Fezzan region. On arrival, both of the countries were concerned with the high levels of illiteracy the country was suffering at that time. Therefore, a large number of schools were constructed in the three regions to allow greater opportunities for the population to receive a satisfactory level of education in various subjects. In their efforts to provide a suitable education in the two regions it governed, Britain adopted the educational system which it had applied in its Egyptian colony. Similarly, in the Fezzan region, France applied the same system it had used in Tunisia (Obeidi, 2001).

1.4.2.2 Current system of Libyan education in Libya

When Libya declared its independence in 1951, the illiteracy rate in the country was 90 %, with only a small number of Libyans holding a university degree (Najeeb, 2012). Since that time however, the country has seen a remarkable increase in the number of educated people, with a literacy rate of about 83% of the total population in 2008 (Najeeb & ELdokali, n. d). Since independence, Libyan governments have been eager to ensure inclusive and good quality education among all people, by providing access to appropriate education for all, both males and females, and in both rural and urban regions. Like other public sectors, the education sector is governed by the Libyan authorities. The Ministry of Education is responsible for planning and making decisions about the education system and curricula. Currently in Libya, there are three main stages of education: basic, intermediate, and higher. Since the early 1970s, education has been free to all members of society from basic to university level, both inside the country and abroad. Pre-university schooling is divided into three phases: primary, preparatory and secondary. The primary and preparatory stages are considered as basic education and are compulsory for all children, whereas the secondary stage is considered as intermediate education and offers a way into higher education. Higher education is comprised of both universities and higher technical and vocational institutions (LNTTO, 2012). Throughout the academic year, students at all of the aforementioned levels attend classes for 6 days per week (Saturday to Thursday). A recent report (Taghavi, 2013) estimates that there are 1,7 million students at all levels, representing about 28% of the total Libyan population. 16% of these students are enrolled in tertiary education, both in universities and technical institutes. This report also showed that since the mid-1970s, the number of students at all levels of education has increased at an annual rate of 2.5%. The order of stages and the anticipated routes that students in Libya may follow throughout their educational journey (including postgraduate education) is presented in Figure 1-1 below. However, because this study is concerned with undergraduate students, a summary of post-graduate education will not be included in the discussion below.

Figure 1-1: The order of stages in the Libyan educational system (adapted from Asker, 2011)



1.4.2.2.1 Basic education

According to the Libyan educational system (GPCE, 2008), basic education includes the first 9 years of school instruction (from 6 to 15 years old). This stage of education is divided into two phases. In the first phase children begin a six-year primary education, starting at the age of 6, and continuing through grades 1 to 6. In the second phase of basic education, at the age of 12, children are enrolled in 3-year preparatory schooling programme (grades 7-9). During the course of their study within these stages, children receive instruction in various subjects including Islamic culture, Arabic language, technical education, mathematics, geography, art, history, natural sciences, and physical education (Gulinao, 2013). The major aim of basic education is to provide pupils with range of knowledge, and help them, at this early stage, to acquire the skills and concepts needed to help them in their further education and to prepare them for lifelong learning. That means that pupils need to learn and pass examinations in a particular number of syllabuses during the 6-year primary educational

programme in order to move on to the next stage; namely the 3 years of preparatory education (Tamtam et al., 2011).

1.4.2.2.2 Intermediate education

The National Report of Libya (GPCE, 2008) states that upon completion of the nine years of basic education, students should be enrolled in a 3-year intermediate program. At this level, students will choose to study one of two types of education: technical and vocational education or secondary education. Technical and vocational education refers to a 3-year programme offered by specific institutes and training centres, which are situated in all parts of the country. These programmes are designed to provide high-quality educational content which aims to improve students' knowledge and skills. At this stage, students study different, specific subjects that are related to industry and also practical technology such as mechanics and electronics. The students who finish their 3-year secondary education in one of the technical and vocational institutes are awarded an intermediate training diploma (GPCE, 2008).

The second kind of intermediate education (secondary schooling) has recently undergone a reform process. According to the former general secondary education system, students could optionally choose between science and arts in order to prepare for university. Students who finished their secondary school having specialised in science would be allowed to join one of the scientific departments at university level (i.e., engineering, chemistry, biology, mathematics, medical science), whereas students who had studied arts would be enrolled in one of the educational or arts university departments (i.e., history, geography, psychology, sociology, and languages). However, due to the recent reform in secondary educational policy, general secondary schools have now been transformed into specialised secondary schools. According to the new system, students can choose from more areas of specialisation in order to select their main subject. Although this specialisation of the secondary school system was approved by the Libyan government in 1992, it was only formally applied

in the whole country in 2004 (GPCE, 2008). A statement given by the Libyan Ministry of Education, (1992, cited in Asker, 2011) proclaimed that, by adopting a system of specialisation at the secondary stage, the educational authority aimed to improve the students' learning and to allow for the identification their specific subject of interest at an earlier stage. The authorities also sought to develop a system that catered for the country's academic and pragmatic needs. In this system there are a varied number of domains of specialisation from which subjects can be selected, including basic sciences, engineering sciences, life sciences, economic sciences, social sciences, and languages. A students' subject of specialisation at the secondary level will determine the academic area they will carry on studying at university level. Table 1-1 below illustrates the enrolment scheme of students in university departments, after gaining their specialised secondary education certificate.

Table 1-1: Enrolment in university education after gaining the specialisation secondary education certificate

	Division	University faculties the students enrol in
1-	Basic Sciences	Sections of Faculty of Science (Maths, Statistics, Physics, Earth sciences, Computer, Meteorology, teacher training colleges and higher vocational training centres)
2-	Engineering Sciences	Various Sections of The Faculty of Engineering, and Teacher Training Colleges and Higher Vocational Training Centres
3-	Life Sciences	Medicine, Dentistry, Pharmacy, Veterinary, Medical technology, Teachers' Training College, Higher Institutes of Health, Faculty of Science Departments (Plant and Animal)
4-	Economic Sciences	Economy, Accounting, Administrative Sciences and The College of Teacher Training and Higher Vocational Training Centres
5-	Social Sciences	Literature, Law, Political Sciences, Physical Education, Arts and Media
6-	Languages	Language Departments and Faculty of Arts and Teacher Training

(source GPCE, 2007, cited in GPCE, 2008)

1.4.2.2.3 **Higher education**

In considering Libyan higher education, it is worth noting that this level of education is very modern in comparison with many other developing countries

(Tamtam et al., 2011). Libyan higher education is governed and financed by the Ministry of Higher Education and is provided by universities and higher technical and vocational institutions. Similar to the other levels of education in Libya, higher education is within the public sector. Private higher education is considered a new-born sector as a law for establishing private universities was only very recently introduced. This law seeks to give the higher education sector a new impetus (Hamdy, 2007). The duration of university study programmes is 4 years and, upon successful completion, students receive a Bachelor's degree. On the other hand, vocational institutes provide 3-year programmes, after which the students are awarded a Higher Diploma in their domain of specialization. While universities enrol students who have completed their secondary specialisations, higher technical and vocational institutions accept students who have gained secondary technical and vocational certificates (GPCE, 2008).

As pointed out above, in Libya, higher education is considered a rather new sector. The establishment of the first university by the authorities, in the city of Benghazi (in the eastern region), in 1955 (shortly after Libya gained independence in 1951), was seen as a major success. When it opened, the University of Benghazi had one faculty (The Faculty of Arts) and a branch in Tripoli. This university later played a central role in establishing other Libyan universities. It was extended between 1962 and 1967, and further new faculties, such as education, sciences, economics and commerce, law, engineering and agriculture were opened in campuses in the cities of both Benghazi and Tripoli (LNTD, 2012). In 1973, the university campuses in these cities were separated and a new university was opened in Tripoli (Tripoli University) with additional faculties being instituted in both of the two universities.

Following that separation, there was a remarkable growth in the number of students who were enrolled in university education, rising from 13,418 students during the 1975/1976 academic year to about 200,000 students in the 2004 academic year (Tamtam et al., 2011). In order to accommodate such rapid

growth, the authorities established new state universities in several different cities in Libya. Currently, there are 8 state universities satisfying the demand for higher education across Libya (Gulinao, 2013). As the enrolment policy illustrated in Table 1-1 shows, faculties in these universities accept students according to their secondary education specialization.

In addition to the subject of specialisation, a particular grade is also required for university enrolment. El-Hawat (2003) explains that according to the criteria established by the Committee of Higher Education in 1990, secondary school students need to pass their final year exams with at least 65% as an overall score in order to be accepted in their chosen department. However, as an exception, study in the faculties of medicine and engineering requires a higher score of 75% or above. If a student finishes with less than 65%, s/he will be accepted in one of the higher vocational and training institutes. As shown in Table 1-1, university faculties consist of a wide range of departments offering a variety of programmes. All teaching staff members in these departments are Doctoral or Master's degree holders. A government report prepared by the Libyan Minister of Higher Education (LMHE, 2010), indicates that university courses used to be taught by overseas instructors and included only a small number of Libyan teachers. However, given the recent policy of promoting postgraduate studies abroad, many Libyans, both male and female, now hold a Doctoral or Master's degree in a wide range of disciplines. Thus recently, more and more Libyan instructors are being assigned to teach in all state universities departments.

Al-Naiely (2012) compared English teaching in Libyan universities and school classrooms and reported a number of differences. Firstly, the Libyan Minister of Education is responsible for managing and designing state curricula for EFL programmes in all schools, whereas university teachers have no fixed curriculum. Instead, university departments provide teachers with broad outlines and teachers individually plan and design their own English syllabuses. Secondly, EFL teachers in schools have to adhere to the teaching of the textbook

and activities decided by the Minister of Education. In contrast, teachers at university level decide and select materials and activities which they feel meet their students' needs. Al-Naiely (2012) also points to another difference at the level of teacher training programmes. While the school EFL teachers receive out of term training programmes offered by the Ministry of Education, which aim to improve their teaching skills, universities do not provide such training courses. Thus, university teachers are left with little chance of utilising the materials and approaches they use in the classroom and reflecting on them in a formal way.

Tamtam et al. (2011) highlight that the Libyan educational authorities have set out a number of objectives for university programmes in order to expand knowledge, enabling the country to achieve economic and social development. One of the major objectives of these higher level courses in English is to improve the status of EFL education in the country. According to Tamtam et al. (2011), the Libyan authorities' current encouragement to the higher education system to include English aims to "keep pace with international standards, the progress made in science, and technology globally [...] ensuring [the] Libyan community becomes fully civilized and admissible in the global community" (p. 747). Along with the establishment of secondary schools that specialise in English, English language departments have been opened in all universities. Before discussing the current status of English language in Libyan universities and its challenges, it is important to provide a brief historical review of the ELT in the country.

1.4.3 Background of ELT in Libya

The earliest stage of EFL teaching in Libya began in 1943, when English was first introduced to the school curriculum in the two Libyan regions (Tripolitania and Cyrenaica) which were under British administration. At that time, Libya was constituted of three regions and French was taught in schools belonging to the Fezzan region, which was administered by France. After gaining its independence in 1951, the Libyan authority sought to disseminate English

language education to the whole nation. At this time, English was first taught through a series of five books called ‘The Modern Readers’, which were presented along with five comprehension volumes. These books were taught in order throughout the 3 years of the preparatory stage and the first two years of the secondary stage (the 5th year of English teaching). Students in the 3rd (final) year of secondary stage were given a syllabus based on English grammar and composition. The Grammar translation method, which involves teaching grammar rules deductively, generally followed by students practising the rules by doing grammar drills and translating sentences to and from the target language, was the common approach to teaching (Shihiba, 2011). It focuses on the learning of vocabulary through reading and translation, emphasising the role of rote learning (Howatt, 1984). Generally, such reading was practised by using simplified novels prepared by members of an inspectorate (UNESCO, 1968).

Since Libya gained its independence, ELT in schools has passed through several stages of development. In the mid-1960s, the Libyan Ministry of Education took steps to improve the status of English language education in the country with the cooperation of UNESCO. Barton, an English language specialist, who carried out the UNESCO mission (UNESCO, 1968), acknowledged that English as a foreign language enjoyed an important status in Libya. In his report, Barton ascribed the growing interest in English education to several reasons, including the government’s attempts to improve the level of higher education in the country, and the rapid industrial and social and economic development.

Together, the UNESCO mission and the Libyan authorities developed an English language syllabus, as part of which, a new series of textbooks called ‘English for Libya’ (first introduced in latest 1960s) was provided to schools. English at that time was taught as a compulsory subject from the preparatory level of education (from the 7th grade of basic education at age of 12). A revised version of this series entitled, ‘Further English for Libya’ was later introduced in 1974. As indicated by Tantani (2012), this material adopted the audio-lingual

approach of language teaching, which focuses on structure and form as well as grammar, and neglects the importance of meaning as an objective of language teaching. The books typically included topics that were related to Libyan culture, followed by reading drills and exercises. By the beginning of 1980s, Grammar-Translation had once again become the common approach for English teaching in Libya. Classroom instruction focused on the principle of teaching grammar as a set of rules, reading aloud, and vocabulary memorisation skills and was characterised by oral drills including correcting learners' errors in grammar pronunciation (Orafi & Borg, 2009). The Grammar-Translation method continued to dominate the classroom of English language teaching until the late 1990s (Emhamed & Krishnan, 2011). However, the breaking point in the history of teaching English language in Libya was in 1986³ when, for political reasons, English teaching was suspended at all levels of education. For nearly a decade, the teaching of foreign languages in schools, including English, was banned (Najeeb & ELdokali, n. d). This disruption resulted in negative consequences in relation to the current status of EFL education, and the level of English proficiency amongst the population (Black, 2007). During the period that English classes were banned in Libyan schools, EFL teachers were assigned to teach other subjects, such as history and political culture or otherwise had to carry out administration work in schools. This long disturbance, as reported by Sawani (2009), affected the teachers' practices and capabilities when English returned to the school syllabus.

In the mid-1990s, the Libyan government decided to resume foreign language teaching in order to meet the demand for the learning of English. Following this decision, Libyan schools continued using their previous materials and Grammar-Translation remained the central method. In this case, the EFL learning process

³ In 1986, foreign language teaching, including English was banned in Libya after US planes bombed Tripoli after a Berlin terrorist attack on US troops was blamed on Gadafy's agents (Black, 2007).

was considered to consist of mechanical habit construction. Teacher-led instruction dominated the classroom learning processes, and largely consisted of asking a series of questions and expecting students to answer them orally. By using the classroom board, teachers may also have demonstrated examples to explain different sets of grammatical rules (Aldabbus, 2008). The Grammar Translation Method was the dominant approach until the late 1990s. However, because there had not been a demonstrable improvement in the status of teaching and learning English in Libyan classrooms, the educational authority realised the importance of revising and developing the existing school EFL syllabus. Therefore, in 2000, the Libyan Ministry of Education designed a new syllabus based on the Communicative Language Teaching approach (Emhamed & Krishnan, 2011). Nevertheless, Aldabbus (2008) suggests that the value of communicative activities is still doubted by some EFL teachers in Libya. He ascribes this fact to the widely held belief that Libyan students are best taught by explicit explanation of a set of lexical and grammatical rules that will help them in learning English language. These views have consequently influenced the classroom practice (Aldabbus, 2008).

As a result of the recent reform of the education system pointed out in section 1.4.2.2, English is now taught as a specialised subject from Secondary school level and the noticeable improvement that ELT in Libya has achieved since the establishment of specialised schools in 2004 is now acknowledged by some EFL teachers. In reviewing the small number of studies conducted in the Libyan EFL context (mainly MA and PhD research), Asker (2011) notes the significant impact of specialised schools on students' levels of competence in English and an overall finding reported by these studies suggests that the majority of EFL students who attended English in specialised secondary schools performed better in their first year at university than students who had completed their secondary education under the old system of general secondary schools.

The following section will shed light on the current status of ELT in Libya with a particular focus on the language teaching and learning at university level.

1.4.4 Current status and challenges of ELT in the Libyan universities

English at university level is a compulsory subject for both English and non-English majors. In common practice, non-English majors (e.g. students in geography, history, Arabic language, and psychology, computer, and biology departments) must attend a two-hour lecture in EAP on a weekly basis. These (EAP) courses are commonly given by English language teachers from the English departments and the syllabuses are flexible, as they are usually planned and designed by the Head of the English department. The students who take these EAP courses are generally taught short scientific texts, which are related to their areas of study. They are required to answer some comprehension questions at the end of the passage with their teacher. They also learn new vocabulary related to their majors.

In contrast, within the English Language Departments, which form the focus of this study, English majors receive more intensive instruction in different EFL skills, including reading, writing, listening, and speaking. In addition to classes on these four skills, EFL students must also take other obligatory modules including, phonetics, applied linguistics, literature, and teaching methods. These modules are taught alongside a very small number of non-English related subjects, namely, Arabic language and Islamic culture. Like students in other university departments, English majors study a four-year English language programme and attend regular classes six days a week. According to Al-Naiely (2012), English-related subjects currently comprise about 80% of the English department curriculum. All programmes are taught by well-educated Libyans or foreign teachers, mostly from India and a few from other Arab and African countries. As pointed out above, there is no fixed curriculum in university departments, hence, EFL teachers select their own materials on the basis of the

syllabus provided by the Head of the department. The teachers are responsible for searching for and determining what they believe to be appropriate for their students, either from within their own materials or from the limited materials provided by their department.

As mentioned above, the Libyan educational authorities have recently called for a shift from the Grammar-translation method to the communicative approach (Emhamed & Krishnan, 2011; Shihiba, 2011). However, in reality, many EFL teachers still choose the Grammar-translation method due to the pressures of examinations, the large number of students in the class and the highly prescriptive course schedules (Saaid, 2010), and university level teaching is no exception. This method dominates the university classrooms, as it fits well with traditional concepts that see activities such as application of grammatical structures, memorization of isolated vocabulary, and translating and understanding reading texts as very important methods for teaching and learning a language. The approach also caters for the highly examination-oriented system in Libyan universities that encourages teachers to prioritise grammatical explanations over communicative activities. In addition, the traditional Libyan methodology highly values the role of memorisation, and this seems to be the most common strategy in teaching of vocabulary and grammar in this country, as teachers are extremely concerned with helping their students to achieve high marks in their English examinations. Therefore the general assumption of most university EFL instructors suggests that in order to understand and be able to communicate in English, students need to primarily master the basic grammatical rules.

In the University EFL classroom, the teacher is regarded as the source of knowledge. S/he is the responsible authority who transmits what s/he knows to her/his learners. In my experience, learners seem more concerned with receiving the language, rather than understanding the meanings or even being able to use it. The learners are often passive receivers and are not expected to challenge the

teacher. Due to this, teachers mainly concentrate on how to pass on their knowledge to learners, instead of being concerned with how individual variables affect learning and how to help learners become more effective learners. In other words, they focus more on the product than the process and so give little consideration to how to help students to learn independently.

Moreover, despite the well-known fact that technology has made a great contribution in terms of providing English language access for learners worldwide, most Libyan EFL learners do not seem to benefit from such an advantage. First, there is an obvious lack of English media in Libya, for instance, only two news satellite channels broadcast in English (BBC and Aljazeera). Moreover, English newspapers are not available in many regions in Libya, though a few are supplied to people in Tripoli (the capital). Furthermore, like many other developing countries, the average internet speed in Libya is very poor and is limited in terms of availability. Many parts of Libya, particularly rural areas, are still lacking internet suppliers or have no internet connection at all. Under the far-reaching influence of such conditions, Libyan university EFL learners commonly experience a lack of variety in their approaches to learning.

1.5 The study: rationale, aims, and research questions

Given the research problem stated in section 1.3 above, it is of interest to investigate the effects of ER on the vocabulary improvement of Libyan university students, since ER is considered to be the most important source of input for developing several L2 skills, including vocabulary knowledge (Nation, 1997). Indeed, this research is driven by two main rationales: the personal motivations of the researcher and the major gaps in the literature of incidental vocabulary acquisition (see Chapter 3). With respect to my personal motivation, working as a lecturer in the Department of English at one of the Libyan universities, I noticed that the students used different approaches to enlarge their vocabulary size. However, one particular case that I found most interesting was

a student who had a much larger vocabulary size than any other student I had taught in the department. She had learnt an enormous amount of English vocabulary. When I asked her about the strategies she had used for learning new words, she simply answered “I read a lot.” After a further discussion with her, I discovered that apart from the texts she read in her reading classes, she also read in English as a hobby. She mentioned that she spent a considerable amount of time almost every day reading English novels bought from abroad by her father, who was a pilot for one of the Libyan Airlines.

The lack of research into the incidental learning of vocabulary in the Libyan context has consequently ‘sparked’ my interest in exploring the possible role that extensive reading may play in improving the vocabulary knowledge of Libyan EFL students. This study was also conducted with the goal of filling the gaps found in the literature of incidental vocabulary research. In reviewing many of the previous studies of reading, I found several serious limitations, many of which were associated with the methodologies used in conducting the studies (see section 3.3). Therefore, with specific reference to current status of teaching and learning English vocabulary in Libya, the aim of the current study is twofold: firstly, the present work aims to investigate the effectiveness of reading on L2 vocabulary development. By conducting an experimental case study of Libyan university EFL majors, I am concerned with the relationship between ER and incidental vocabulary acquisition. Secondly, the study also aims to tackle the major methodological problems, which have devalued findings from previous incidental vocabulary research, by developing a novel experimental design. The research questions to be answered in this study are:

1) Does reading extensively enhance adult Libyan EFL learners’ vocabulary knowledge?

To what extent can a two-month ER programme enhance the incidental learning of English vocabulary items by Libyan EFL learners for both receptive and productive use?

- 2) Does word repetition affect receptive and productive incidental vocabulary learning from ER? If so, how many exposures are needed for incidental vocabulary learning?
- 3) How much of the newly learnt vocabulary is retained 9 months after the reading programme?
- 4) Will individualisation in the selection of ER texts reveal any significant difference in the number of vocabulary items learned, as well as in the learners' perception of the ER programme?

1.6 Significance of the study

There are a number of elements which may contribute to the significance of the present study. However, the main area of significance is its aim to explore a substantial trend in EFL instruction that has not yet received sufficient attention. Despite the successful research and growing interest in the role of ER in L2 learning, to the researcher's knowledge, no single study has investigated vocabulary development through ER with Libyan EFL learners. Thus, the current study will fill a significant gap which exists in the research, by evaluating the use of ER programmes in Libyan university contexts and measuring its effect on incidental acquisition of L2 vocabulary.

The prime goal of the study therefore, is that along with the Libyan Ministry of High Education's current efforts, I wish to improve the state of English language teaching and learning in Libyan universities. Findings from this experimental work may raise the awareness of policy makers of the significance of ER for improving the lexical knowledge of EFL learners. The present situation of teaching and learning in EFL classrooms may also be improved by integrating ER into the curriculum. Enhancement of reading practice by providing ER materials is believed to offer greater English language input, which could increase learners' exposure to a large number of new words and facilitate their

incidental acquisition. Carbery and Yoshida (2003) state that for EFL learners to obtain a maximum outcome from reading, an eclectic approach, involving intensive and extensive reading, should be adopted. The study is also intended to spark the interest of other Libyan researchers to carry out other similar studies, perhaps within other levels of education, which may further investigate the incidental acquisition of vocabulary at other levels of knowledge.

1.7 Structure of the thesis

This thesis consists of eight chapters which are outlined as follows:

Chapter one is an introductory chapter. It introduces the background to the study. It also states the research problem and highlights the significance, rationale, and aim of this experimental work as well as the context of the study. The Libyan educational system is also briefly presented followed by background information concerning EFL education in Libya including its current status and the challenges it faces.

Chapter two reviews the relevant literature. It first sheds light on the area of L2 vocabulary acquisition including aspects that constitute word knowledge, particularly the levels of receptive and productive knowledge of vocabulary, as well as the difference between the two approaches to lexical learning: incidental and intentional. Theories of reading, as well as the ER approach, will also be discussed in chapter two.

Chapter three reviews the key studies that have investigated the incidental vocabulary learning from reading and ER in various worldwide EFL and ESL contexts and the limitations that have devalued these studies. This will be preceded by a small scale presentation of original studies which have been conducted in L1 contexts.

Chapter four explores the methodology of the study and the underpinning philosophy behind the choice of methodology. It also provides an explanation of the procedures involved in the pilot study.

Chapter five gives a detailed description of the research design used in this study. It starts with provision of information about the research site and gives informative descriptions of the sampling procedures and the samples from the research. It also explains the procedures and instruments used in carrying out this experimental research, and describes the different stages of data collection.

Chapter six answers three of the research questions posed by the study. This entails statistical analyses of the data and a discussion of the main findings.

Chapter seven proceeds to answer the fourth research question by analysing the individualised statistical data, as well as qualitative data collected through interviews with the participants, and discusses the findings.

Chapter eight provides the conclusions, contributions to knowledge, limitations of the study and some possible suggestions for future research. The chapter also presents the pedagogical and empirical implications of the present study.

Chapter 2:

Review of literature

Chapter 2: Review of literature

2.1 Introduction

This chapter offers an integrated review of the literature relevant to this study. This review of literature is divided into two major themes: L2 vocabulary acquisition, and definitions and theories of L2 reading. In order to obtain knowledge about these issues, a large number of the existing publications in the SLA area will be reviewed. Important aspects related to word knowledge and incidental and intentional learning of vocabulary will be discussed first. The second part of the chapter will shed light on theories of reading and some of the Arab learner's problems with EFL reading. It will also provide a review of the literature pertinent to ER, including its underpinning theory, principles and its advantages and drawbacks.

2.2 L2 vocabulary acquisition

For decades, vocabulary acquisition has been a neglected area in the field of L2 education. Oxford and Scarcella (1994, p. 241) highlight this point, by stating that “vocabulary instruction is ignored in many L2 classes, on the assumption that students themselves will simply find ways to memorise words without any help”. Under these conditions, it is not surprising to find that only a small amount of time is given to vocabulary teaching, and the teacher's focus in many L2 classrooms around the world has been on grammatical patterns and sentence structure. However, after decades of neglect, vocabulary learning has recently gained considerable interest within the field of second language acquisition (Carter & McCarthy, 1988; Lewis, 1993; Coady & Huckin, 1997; Schmitt & McCarthy, 1997; Schmitt & Meara, 1997; Schmitt, 2000; Nation, 2001; Milton, 2009; Schmitt, 2010).

Krashen (1989a) suggests that language learners cannot achieve language mastery without learning a large number of words. Similarly, in my experience as an EFL teacher in Libya, many language learners seem to be well aware that their lack of vocabulary knowledge is their major problem with learning. More recently, during the 1980s and 1990s, the L2 arena has witnessed an upsurge of research into vocabulary acquisition. However, the description of the process by which vocabulary is mentally acquired has been a problematic issue for SLA researchers. Schmitt (2000) ascribes this problem to the fact that researchers cannot view the words contained in a learner's mental lexicon, and this will remain true until neurologists can physically find and follow words in the brain.

Regardless of the lack of theoretical knowledge of how vocabulary is acquired, a number of researchers have undertaken empirical studies which have endeavoured to describe the process by which second language learners might acquire different word features, (For example, Mondria & Wit-De Boer, 1991; Wang et al., 1993; Coady & Huckin, 1997; Yang, 1997; Hulstijn & Laufer, 2001; Laufer & Rozovski-Roitblat, 2011). The consensus in SLA literature is that vocabulary may be acquired in different ways, but the most debated distinction is the one made between incidental and intentional learning, or between implicit and explicit learning. Explicit and intentional learning are generally characterised by the level of consciousness involved in the learning process (Schmidt, 1990), but both of these learning modes involve the study of words in isolation, and the use of dictionaries and glossaries (Hulstijn, 1992). The lack of clear operational definitions for incidental and implicit learning, on the other hand, makes the distinction between the two approaches particularly complex to explain. Because of its relevance to the present work, previous research about incidental vocabulary learning will be discussed later in Chapter 3. However, prior to that, some theoretical concepts regarding word knowledge and incidental and intentional learning of L2 vocabulary will be considered.

2.2.1 What constitutes word knowledge?

For many second L2 learners, knowing a word means recognising its meaning, as well as its written and spoken forms (Nation, 2010). However, learning a word seems to be more complicated than simply making a connection between the word's meaning and form (Laufer & Goldstein, 2004). Pignot-Shahov (2012) suggest two reasons why such a view is insufficient for describing the concept of word knowledge. First, it indicates that meaning and form of a word are two separate aspects of word knowledge. Second, word acquisition is a complex process involving knowledge of many other aspects of a word that this definition does not mention.

With this in mind, different researchers have suggested different taxonomies for word knowledge. The earliest framework of word knowledge was proposed by Richards (1976). According to him, learning a word involves several aspects of its knowledge. These aspects include knowing:

- the degree of probability of encountering the word in speech or print,
- the limitations imposed on the use of the word according to the function and situation,
- the syntactic behaviour associated with the word,
- the underlying form of a word and the derivations that can be made from it,
- the associations between the word and other words in the language,
- the semantic value of the word, and
- the different meanings associated with the word (p.83).

Description of word knowledge in many vocabulary studies has been directed by these assumptions. However, this list is not exhaustive and has received criticism from Qian (2002) for not including the aspects of pronunciation and spelling. In a similar vein, Meara (1996) points out that that Richards' (1976) framework does not appear to offer a suitable theoretical model for defining

word knowledge, but instead could be more suitable for guiding classroom instruction.

Later on, Nation (1990) elaborated on and extended Richards' (1976) taxonomy by suggesting eight types of word knowledge. For Nation, knowing a word entails knowledge of: a) its spoken form, b) its written form, c) its grammatical behaviour, d) its collocations, e) its frequency of use in a language, f) its register (*knowing the constraints which determine if a word is appropriate in a particular context*), g) its conceptual meaning, and h) its semantic network of association.

According to Henriksen's (1999) taxonomy, vocabulary knowledge is an even more complex construct. She proposes a three-dimensional framework for knowing a word. The first dimension is partial vs. precise knowledge, which suggests that words are known to a certain degree, but certain aspects of a word may not be fully known in every case. The second aspect is a depth of knowledge dimension, which refers to network building. Words are believed to be stored in networks in the brain (Haastrup & Henriksen, 2000), and the stronger and the greater the number of links between the words (antonyms, synonyms, hyponyms, gradation and collocations), the more developed the learners' vocabulary knowledge becomes. These two dimensions are related to the comprehension of word knowledge. The final dimension deals with receptive and productive vocabulary use and is related to the access and control of the comprehension of word knowledge.

Nation's (2001) classification, which is perhaps the most comprehensive model of word knowledge up to date, differs slightly from Henriksen's (Webb, 2007). As illustrated in Table 2-1, Nation uses the terms *receptive* and *productive* to describe word knowledge, covering all aspects of what is involved in knowing a word. Like Henriksen (1999), Nation (2001) also subcategorises lexical knowledge into knowledge of word *form*, *meaning*, and *use*. However, Nation

provides a more detailed description of the receptive and productive aspects contained in each of these three dimensions of word knowledge.

Table 2-1: Aspects involved in word knowledge

Form	spoken form	R	What does the word sound like?
		P	How is the word pronounced?
	written form	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 the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	what words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (register, frequency)	R	Where, when, and how often would we expect to meet this word?
		P	where, when, and how often can we use this word?
R = receptive knowledge		P = productive knowledge	

(Adapted from Nation, 2001, P. 27)

According to Nation’s classification in Table 2-1, knowing the *form of a word* requires knowledge of its spoken and written form as well as its parts, whereas knowing word’s *meaning* involves knowledge of the link between its form and meaning, and the association between the concept of meaning and referents. Knowledge of word *use* requires knowledge of its grammatical functions, collocations, and the constraints on its use. Other aspects such as knowledge of multi-word units are not included in any of these taxonomies, despite being considered part of word knowledge. A justification for that is found in Laufer and Nation (2012, p. 170), who acknowledge that “we [researchers] do not yet have standardised tests that measure global knowledge of these units”.

However, there is a consensus in all of the models of word knowledge reviewed above that vocabulary knowledge is multifaceted and comprises a

range of aspects. Disappointingly, only some of these aspects, such as word meaning and form, have captured the attention of researchers in the L2 vocabulary acquisition arena, whereas other major types of word knowledge such as word use (including grammatical functions) have rarely been investigated (Waring & Takaki, 2003; Pigada & Schmitt, 2006; Webb, 2007). Similarly, there has been little research into how receptive and productive vocabulary knowledge is learned (Webb, 2005). In the present study, incidental knowledge of new vocabulary will be defined in respect of its receptive and productive use. That is, the study will examine the development of the receptive knowledge of word meaning from a written context and the productive knowledge of its use in a written sentence.

2.2.2 Receptive and productive word knowledge

One well-established convention in the field of vocabulary research is the division of word knowledge into *receptive* or *passive* knowledge and *productive* or *active* knowledge (Milton, 2009). According to the definition of these two terms, receptive knowledge of vocabulary refers to the learner's ability to perceive a form while listening or reading and to retrieve its meaning. Conversely, productive use of word involves the act of expressing a meaning through speech or writing and the ability to retrieve or/and produce an appropriate spoken or written form of a word (Nation, 2001; Milton, 2009). It is generally acknowledged that language learners' vocabulary receptive knowledge is likely to be greater than their productive knowledge. Some estimates suggest that 50-80% of words a learner has acquired receptively are also known productively (Milton, 2009). However, an issue which has witnessed a large debate is whether the learning of vocabulary for receptive and productive use is "dichotomous" or whether "it constitutes a continuum" (Laufer & Goldstein, 2004, p. 405). A very common theoretical assumption is that vocabulary knowledge is acquired in continuous terms, in that, language learners acquire receptive knowledge first and then, through further exposure to a word, their knowledge of it gradually improves to a productive level of

use (Melka Teichroew, 1982; Faerch et al., 1984). However, the distinction between receptive and productive knowledge is more complex than it appears in this definition. Schmitt (2010, p. 82) argues that due to the complicated nature of word knowledge, language learners do not acquire receptive and productive facets in “a uniform manner”. Schmitt (2000) also refers to the current lack of theoretical understanding about the threshold at which receptive knowledge moves to a productive level. As to what constitutes word knowledge in the above discussion, it is clear that all researchers (Richards, 1976; Henriksen, 1999; Nation, 2001) have found it difficult to define the terms receptive and productive, and consequently, they have used “the same terms [...] for different processes or subprocesses” (Pignot-Shahov, 2012, p. 83).

With respect to ESL/EFL research on incidental vocabulary learning from ER, interest has been largely focused on assessment of the size of learners’ receptive gains. Most of the studies asked subjects to choose an appropriate meaning from a number of possible choices or to provide synonyms or a translation to show that some knowledge of form and meaning has been acquired (for example, Horst, et al., 1989; Pitts et al., 1989, Day et al., 1991; Grabe & Stoller, 1997; Waring & Takaki, 2003). Nonetheless, Yamamoto (2011) noticed that very little research has been carried out to explore whether ER is responsible for increases in productive vocabulary knowledge. The lack of a more comprehensive assessment of vocabulary gains is probably due to the absence of appropriate tests which can comprehensively measure every facet of vocabulary knowledge (Pignot-Shahov, 2012). This complex issue is raised by Milton (2009, p. 117), who says:

[m]easuring the productive vocabulary that learners possess poses methodological problems for the investigator in how best to capture this quality. The problem is not so much how to devise a test, but how to choose from the many approaches that the researchers have used. A single, definitive method of measuring this quality of knowledge has yet to emerge.

Unlike most previous L2 ER research, this study makes an attempt to assess incidental learning at both receptive and productive levels. Since it cannot be expected that the participants will gain productive knowledge of all vocabulary they are exposed to, the present study bridges the gap between word receptive and productive acquisition by using scales of knowledge testing (by Schmitt & Zimmerman, 2002). Similar to some previous reading research, receptive word learning in this study is measured in terms of the form and meaning relationship. In addition to assessing the participants' word learning rate at a receptive level, they were also required to provide an equivalent translation for a word. This study, however, extends the word knowledge assessment and looks at the productive level by testing the learners' ability to produce sentences using the newly acquired lexical items from the extensive reading. In the following section, some discussion of the incidental and intentional approaches will be presented. This includes the advantages and drawbacks of each approach.

2.2.3 Incidental vs. intentional vocabulary learning

Learning vocabulary incidentally from context is usually seen as being in opposition to the explicit, intentional learning of words (Nation, 2001). However, the lack of a strong theoretical explanation for incidental learning appears to cause confusion when trying to define it, although attempts have been made by several authors to describe its process (Schmidt, 1990; Ellis, 1994; Schmidt, 1994; Huckin & Coady, 1999; Kerka, 2000; Richards & Schmidt, 2002). The key aspect in defining the term 'incidental' is closely linked to the notion of 'attention'. In operational terms, the distinction between incidental and intentional learning relies on whether or not the learners are aware of the intended implementation and purpose of a post-reading vocabulary test (Hulstijn, 2001). Wesche and Paribakht (1999) point out that "the critical factor distinguishing incidental from intentional learning conditions is that subjects in the 'incidental condition' are not told they will be

subsequently tested on given material or skills, whereas, in the ‘intentional condition’, they are told” (p. 176).

Even though the distinction between the two learning modes is relatively clear-cut in operational terms, cognitive psychologists seem to have found difficulties in reaching a consensus on conceptual definitions for the two terms. This debate, as noted by Hulstijn (2001), is caused by the complicated nature of the processes involved in these two learning conditions. The common definition for incidental learning, however, is learning one thing while intending to learn something else (Richards & Schmidt, 2002). More specifically, Ellis (1997, p. 35) defines it as the learners’ ability to acquire L2 words and rules, whilst they are focused on trying to understand and communicate using language which does not necessarily require ‘conscious intention’. For instance, this type of learning can occur while the learner is engaged in different communicative tasks, such as reading and listening (Ellis, 1999; Hulstijn, 2003). Kerka (2000) describes incidental acquisition as unintentional or unplanned learning, although Milton (2008) argues that these two terms refer to two different concepts. While unintentional learning suggests that skills can be acquired without really trying and without effort, presumably when the learner is engaged in other activities, unplanned learning may involve “the deliberate intention to learn even if the activity involved is not part of a formal syllabus or curriculum; listening to a song and trying to memorise the words, for example” (2008, p. 229). Intentional learning, in contrast, is defined as type of learning in which attention is directly focused on the information being learnt. It is always designed, planned for, or intended by teachers or students and involves some explicit activities focused on practice and learning (Schmidt, 1990; Ellis, 1999; Gass, 1999; Schmitt, 2000; Hulstijn, 2001, 2003). It also includes any efforts and techniques involved in the intentional nature of the action (Gass, 1999, p. 321).

This definition of ‘incidental’ language learning, seems congruent with Krashen’s (1989a, b) IH, which claims that language is acquired in one way only; by understanding the message in a discourse by means of comprehensive input and incidental learning (Krashen terms it ‘acquisition’). Therefore, acquisition occurs when the attention of language learner is focused on understanding the meaning of a message. Several authors ascribe these processes of incidental acquisition as a ‘by-product’, not the target, of other cognitive activities that are involved in comprehension, such as reading (Gass, 1999; Huckin & Coady, 1999; Wesche & Paribakht, 1999; Wode, 1999). This incidental acquisition typically occurs when the word is repeatedly encountered in various comprehensible contexts.

However, the description of incidental learning as a ‘by-product’ of other cognitive activities has been criticized by other scholars, as it appears to ignore any active role the learner may play in the learning process. It has also been argued that attention to unfamiliar vocabulary is essential for any incidental acquisition to occur (Gass, 1988; Ellis, 1994; Schmidt, 1994). If consciousness is equated to intentionality, Ellis (1994, p. 38) stresses that as a term, “incidental learning should be used to refer to describe situations where individuals learn without intent to learn, or when individuals learn one thing when their primary objective was to do something else”.

The distinction between incidental and intentional learning is described by Ellis (1999) as the difference between ‘*peripheral*’ and ‘*focal*’ attention. He argues that in the incidental learning condition, a learner gives their main attention to the meaning of the message, but some partial attention is allowed for a focus on the form. In the intentional learning condition on the other hand, the learner’s focal attention is purposely placed on the language and in particular on making an association between form and meaning. This distinction is also echoed by Schmidt (1990) who claims that a certain degree of consciousness is involved in the incidental acquisition process that happens

when a learner ‘notices’ a new lexical item or rule in the input. Thus, learning, whether it is intentional or incidental, cannot operate without a degree of attention.

Although, terminologically, incidental and intentional are often used in connection to the implicit and explicit conditions of language learning, the meaning of the two learning dichotomies appears to overlap in the SLA literature. The terms incidental and implicit have been commonly used side by side to describe the way in which ‘picking up’ of new knowledge may take place (Rieder, 2003). The distinction between the two learning processes is apparently based on the amount of consciousness involved in the learning process. Accordingly, implicit learning is seen as “acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, simply and without conscious operation” (García, 1991, p. 1), or without consciously knowing what is being learnt (DeKeyser, 2003). Incidental learning, on the other hand, is typically characterised as involving some level of conscious attention (Schmidt, 1990). Even with this attempt to distinguish between incidental and implicit learning, it still seems difficult to determine which learning is incidental and which is implicit. In his comments, Rieder (2003) states that in principle, this terminological confusion, appears to be a result of the difficulty of providing a precise interpretation of the term consciousness, or what Ellis (1994, p. 38) calls the ‘slipperiness’ of the term ‘consciousness’.

In the domains of both L1 and L2 acquisition, incidental acquisition is considered to be the most effective approach for vocabulary learning (Hunt & Beglar, 1998). Therefore, it is generally agreed that incidental learning accounts for some, if not a large part of acquired vocabulary and only a small amount results from direct (intentional) instruction (Saragi et al., 1978; Nagy & Herman, 1985; Nagy et al., 1987; Day et al., 1991; DeKeyser, 2003). It has been estimated that, with the exception of the first few thousand most frequent

words in a language, vocabulary acquisition largely takes place through guessing the meaning of unknown words while the learner is involved in ER (Huckin & Coady, 1999).

Therefore, for L2 acquisition, the approach seems to have significant pedagogical importance, particularly for learners who are required to acquire a large L2 vocabulary in order to enable them to use the language efficiently. In fact, it is estimated that English native-speaking university graduates know around 20,000 word families (Goulden et al., 1990). However, it is obvious that with the very limited time available for L2 direct vocabulary instruction, this number is beyond most L2 learners. The only possible way L2 learners may attain a native vocabulary size is by incidental learning, through activities which involve extensive language input.

While the importance of incidental vocabulary is acknowledged in the literature, certain explicit vocabulary teaching seems to be an essential approach. Schmitt (2000) points out that L1 incidental word learning from listening occurs from a very early stage of learning. However, some explicit vocabulary teaching might be required as a prerequisite for incidental acquisition from reading. This is because spoken language usually uses the *most* commonly used words in a language, while less frequent vocabulary is typically used in written language (Schmitt, 2000). This need for explicit teaching can also apply to an L2 context. In fact, it has been suggested that certain explicit learning is necessary for L2 learners to reach a vocabulary ‘threshold’ which enables the incidental acquisition of words from reading (Schmitt, 2000). One estimation by Nation and Waring (1997) suggests that a vocabulary size of three to five thousand word families is the threshold which enables an adult native learner to begin reading authentic materials. This size seems very high for a L2 learner who is reading simplified texts. Furthermore, for intermediate and advanced learners who have 3.000 words or above, it is proposed that the activity of reading can provide access to the remaining words

(Schmitt, 2000). This suggests that in order to help L2 learners to become fluent readers, explicit vocabulary teaching and learning must focus on the most frequent word families before the learners can be involved independently in extensive reading. Grabe and Stoller (2002) recommend that as an essential foundation for automaticity of word-recognition, any explicit word activity should focus on the second and third thousand most frequent words in a language.

A comparison between incidental and intentional learning indicates that incidental vocabulary has certain advantages over explicit teaching of new vocabulary. One major benefit is that successful guessing of meaning of a word or phrase from its context can lead to more effective long-term retention (Hulstijn, 1992). The underpinning assumption is that some strategies, including guessing word meaning from context, usually involve 'deep processing of information' (Craik & Tulving, 1975), which is believed to enhance retention. In contrast, direct de-contextualised learning of vocabulary items, such as from a word list, is likely to result in a rapid rate of forgetting (Mondria & Wit-De Boer, 1991; Oxford & Scarcella, 1994). This is because while a language learner is involved in the guessing process, s/he is performing a mental process concerning word form, attempting to link context to his/her linguistic and personal knowledge. All of these processes lead to a cognitive foothold (Schouten-van Parreren, 1985, cited in Mondria & Wit-De Boer, 1991).

The importance of incidental vocabulary learning from ER is justifiable in different ways. Huckin and Coady (1999) suggest three benefits of incidental vocabulary learning from reading. Firstly, this type of learning is contextualised, which gives the learner a richer sense of a word's use and meaning than can be offered in traditional explicit exercises such as word lists, word association and synonym puzzles. Moreover, a comparison of learning through these explicit activities with the inference of meaning of

unknown words through context suggests that the latter is more productive because learners are engaged in cognitive processing, whilst attempting to infer the meaning of unfamiliar words. In addition, the contextualised learning of words sharpens the learners' ability to guess, which is an essential skill for vocabulary learning (Jacoby et al., 1979). Secondly, it is a "pedagogically efficient" approach to L2 learning because it enables the two activities of reading and vocabulary learning to take place at the same time (Huckin & Coady, 1999, p. 189). Thirdly, in comparison to the explicit mode of learning, incidental learning is more individualised and learner-based since the words being acquired rely on the learner's own choice of reading materials.

Despite the crucial part played by incidental acquisition from reading in expanding language learners' vocabulary, the approach is said to have a number of limitations. One of the arguments being advanced is that it can be a slow process; especially, when there are long gaps between word encounters (Webb, 2008a). Therefore, L2 learners may need to spend large amounts of time on reading in order to incidentally acquire a substantial vocabulary size (Carter & McCarthy, 1988). Furthermore, research by Nagy and colleagues (Nagy et al., 1985) has suggested that only 5 to 15% of incidental vocabulary is picked up from reading, but listening accounts for the remaining percentage of incidental acquisition. Another argument is that L2 learners, particularly those at low proficiency levels, are very likely to guess incorrectly (Hulstijn, 1992; Sökman, 1997), so incidental acquisition does not necessarily result in learning the correct meaning of words.

Moreover, guessing unknown words from a context may not be effective if learners' comprehension is not adequate. Several studies have found that the ratio of known to unknown words in a text is an essential factor for inferring the meaning of unfamiliar words successfully (Shelfbine, 1990; García, 1991; Laufer, 1997; Paribakht & Wesche, 1999). Hu and Nation (2000) suggest that for successful guessing, learners needed to have between 96 to 99% knowledge

of vocabulary in a text because the more clues there are, the easier the guessing will be. Nevertheless, there is a counter argument which claims that not all guessed words are necessarily learned from reading. Learners may apply the minimum amount of mental processing when words are easy to guess in a very rich context. This low level processing therefore, may not lead to word retention (Schmitt, 2000).

In conclusion, despite the established fact that vocabulary can be acquired incidentally when learners are involved in language learning activities (i.e., reading and listening), some deliberate and intentional focus seems to be necessary in order to develop skills and strategies such as guessing, which will enable the learners to carry out the activity successfully (Nation, 2001). Moreover, it should be noted here that using incidental learning as a central approach for learning L2 vocabulary is not recommended by many scholars, as it does not appear to be a practical method by itself. Instead, it should actually be supplemented by some intentional learning. As a consequence, both learning approaches should be considered as enhancing the other (Nation, 2001).

Because this study investigates incidental vocabulary from ER, it is felt important to offer some theoretical background into how learners process language during reading in order to gain comprehension of printed texts. The next section discusses the most common reading models in the literature of L2 language.

2.3 L2 Reading theories

Before offering some theoretical insights into reading process, it is important to explore some of the definitions.

2.3.1 Defining reading

Reading is defined as “the process of construction of meaning through the dynamic interaction between the reader’s existing knowledge, the information suggested by written language, and the context of the reading situation” (Anthony et al., 1993, p. 284). Goodman (1967, 1988) defines the reading process as ‘a psycholinguistic guessing game’ in which “[t]he writer encodes thought as language and the reader encodes language to thought” (Goodman, 1988, p. 12). Similarly, Day and Bamford (1998) describe the process of reading as construction of meaning from a written message. In order to arrive at an understanding of the content of a text, the reader connects his/her previous knowledge to the information in the text. More recently, a definition by Grabe and Stoller (2001) suggests that reading refers to the learner’s ability to draw meaning from a printed text and interpret this information appropriately.

The general picture which can be drawn from these definitions is that reading involves various forms of activities, which are described by Pulido (2004) as a complex cognitive process of construction and integration of information. Researchers seem to agree that the relationship between language and thought is a key component of the reading process. Goodman (1988) claims that reading becomes more effective if the reader’s understanding of meaning reaches a certain degree of agreement with the message conveyed in the text. According to Goodman, one of the most important features required for reading efficiently is that the readers must apply the least amount of effort. This, as Goodman asserts, can be attained when readers (a) “maintain constant focus on constructing the meaning throughout the process, (b) always seek “the most direct path to meaning”, (c) always apply “strategies for reducing uncertainty” and (d) are always “selective about the use of the cues available and draw... deeply on prior conceptual and linguistic competence” (Goodman, 1988, p. 12).

Reading, therefore, has been viewed from different viewpoints: cognitive, sociocultural, and physiological (Day & Bamford, 1998). This study, however, endorses a cognitive view of the reading process, since the cognitive perspective offers support to the ER approach to acquisition (Day & Bamford, 1998). Some suggested theories of L2 reading will be discussed in the following section.

2.3.2 Models of L2 reading processes

Over the last four decades, reading researchers have been interested in understanding and describing the processes a reader utilises to extract meaning from a text. Grabe (1991) mentions that most of the current views of L2 reading are shaped by research on L1 learners, as it has a longer history. He therefore, comments that:

[i]t makes good sense, then, for second language researchers and teachers to consider what first language research has to say about the nature of the fluent reading process and the development of reading abilities. A primary goal for ESL reading theory and instruction is to understand what fluent L1 readers do, then decide how best to move ESL students in that developmental direction (1991, p. 378).

Generally, three popular approaches to understanding reading have been developed by a number of psychologists: the bottom-up, top down, and interactive models. These models will be discussed in this section.

2.3.2.1 Bottom-up processing in reading

Researchers who uphold the bottom-up account view the reading process as ‘text-driven’ (Gough, 1972; Rayner & Pollatsek, 1989). As Martinez-Lage (1995, p. 70) states, the printed page is seen as a “chain of isolated words, each of which is to be described individually”, and the reader is the one “who approaches the text by concentrating exclusively on the combination of letters and words in a purely linear manner”(p. 70). The Bottom-up model, therefore, suggests that the meaning of the text is contained only in the text and the reader

takes meaning from a printed text itself (Zainal, 2003). According to this view, reading progresses from lower-level processes, such as recognising the letters of a word, up through the higher-level processes of understanding the meaning of the whole text (Bernhardt, 2011). Hedgcock and Ferris (2009, p. 17) claim that “reading is initiated at the “bottom” level of text structure, from discrete, visual units, such as graphemes, morphemes, and words”, and in order to construct the meaning from the printed page, “ the reader works his/her way “upward” to larger-level units such as phrases, sentences, paragraphs, and chunks of written discourse”.

Strong proponents of bottom-up processing have proposed a number of models, which all suggest that learners deal with letters and words in a systematic manner. Gough (1972) offers a popular bottom-up processing model of reading, which assumes that reading is initiated with the process of finding a fixation point. This fixation point enters the visual system in the ‘iconic memory’, where it remains available until another fixation is made by the reader. It is suggested that letters are the starting point of the reading process; therefore the idea proposed by this model is that the reader obtains information in a sentence by understanding it letter-by-letter, word-by-word and that the words of the sentence are understood in sequence, from left to right.

LaBerge and Samuels (1974) suggest a model of automatic information processing which is built on the concept of attention. In this theory, it is assumed that automaticity of word recognition is a prerequisite of comprehension. The theory argues that for the reader to derive meaning from a written text, two essential steps must be involved, firstly written words must be decoded, and secondly decoded words must be comprehended. This model claims that a fluent reader is able to decode and comprehend printed texts without requiring conscious effort (LaBerge & Samuels, 1974). The process of automaticity suggests that fluent reading must occur before comprehension is possible. In other words, a reader’s ability to decode words determines the

extent to which the words can be comprehended. This is not a straightforward process however, as successful comprehension is more likely possible when accurate and automatic decoding is involved (LaBerge & Samuels, 1974; Eskey, 1988).

More importantly, this model suggests that there is a direct link between attention and reading fluency. For this reason, LaBerge and Samuels (1974) incorporate the concept of internal attention, external attention, and automaticity into their model of reading. According to this model, external attention is related to learner's efforts to use his/ her sensory organs (eyes) to maximise information input. Internal attention, on the other hand, is related to the learner's ability to choose their focus on a particular subject. It is proposed that, in the early stages of learning, the reader switches her/his attention back and forward between decoding and comprehension. However, at a more advanced stage, when reading becomes more automatic and fluent, little or no attention will be required. Attention is exclusively directed to comprehension, when decoding letters of words becomes automatic and independent of attention (Samuels & Flor, 1997) . LaBerge and Samuels (1974, p. 295) argue that "we can only attend to one thing at a time, but we may be able to process many things at a time so long as no more than one requires attention". According to this assumption, beginner readers are slower than the fluent readers because the processes of switching the attention between decoding and comprehension of words in the text takes more time than when the attention is devoted entirely to understanding the text (LaBerge & Samuels, 1974).

Despite the fact that these bottom-up models have been widely cited in the literature of reading, they suffer from several limitations. For instance, the model proposed by Gough (1972), "does not really deal with how higher-order comprehension processes (such as integrating sentences and propositions together) takes place" (Rayner & Pollatsek, 1989, p. 467). In addition, little is said about how sentences are processed, and how the reader derives the

meaning of texts from letters and sentences. Moreover, the model equates reading with speech and does not provide a well-defined mechanism of how information is processed at different stages (Hedgcock & Ferris, 2009). Likewise, the automaticity model described by LaBerge and Samuels (1974) has also undergone criticism for its lack of clear explanation of how *meaning* is important in reading. Bernhardt (1991) points out that since the automaticity model stresses attention and perception as two central components of the process of reading, comprehension of the text becomes a marginal and unimportant element. A further problem with this model is addressed by Rumelhart (1990), who argues that LaBerge and Samuels fail to describe the effect of syntax on the ‘semantic memory system’ they propose in their model of information processing. In Rumelhart’s words, while it is easy to see how this model of information processing “could account for effects of orthographic structure on letter perception, it is somewhat more difficult to see how the effects of syntax and semantics can be mediated with such a model” (1990, p. 582).

These suggested limitations of Bottom-up models have led to the emergence of other theoretical assumptions, which have tried to provide a clearer picture of the processes that take place within reading. Top-down theory, another model of reading, will be discussed next.

2.3.2.2 Top-down processing in reading

In contrast to bottom-up processes, the uptake of information in the top-down model is directed by the reader’s prior knowledge and expectation. Barnett (1989, p.13) defines top-down models as ‘reader-driven’ models where “the reading process is driven by the reader’s mind at work on the text”. Theories which stress the top-down processing of information claim that the reader first creates hypotheses about the conceptual meaning of a text, and then uses the least amount of information necessary from the text to confirm or deny the hypotheses (Urquhart & Weir, 1998). Unlike bottom-up models, which

emphasise that the reader builds up information from the printed letters on the page, top-down processing operates on the principle that reading moves from top “higher level” mental stages, and continues through several steps down to the “lower level” stages, concentrating on the text itself, in which the semantic and phonological features of a language are involved (Barnett, 1989). Nuttall (1996) uses the concept of the eagle’s eye view to the landscape to describe the top-down processing approach. She suggests that the eagle can see a wide area spread out below when it is at a great height. From that high location, the eagle “understands the nature of the whole terrain, its general pattern and the relationship between various parts of it, far better than an observer on the ground” (1996, p. 16).

One of the most well-known top-down models is that offered by Smith (1971), who describes reading as a psycholinguistic process. Smith’s top-down model considers reading as a text-driven process and the reader’s use of the contextual information available in a text plays a facilitating role in understanding its meaning. In this sense, the model hypothesises that good readers are “sensitive to redundancy afforded by sentences”, which allows them to develop hypotheses about the upcoming words and confirm their guesses. This aspect is also reflected by Oakhill & Beard (1999, p. 15), who note that “[g]ood readers should process words faster because their use of redundancy lightens the load for their word decoding mechanisms.

Goodman, another top-down theory advocate, describes reading as “psycholinguistic guessing game ... [which]... involves interaction between thoughts and language” (1976, p. 498), or, more simply, "a dialogue between the reader and the text" (Grabe, 1988, p. 56). Goodman’s (1967, 1976, 1988) top-down model neglects grapheme-phoneme knowledge as the most important source of knowledge. Instead, it asserts that reading is an active process between the reader, writer and the text, of building up meaning. The model also considers reading as ‘hypothesis verification’, whereby a reader makes

hypotheses about the upcoming text while reading. The reader samples as little from the printed page as is needed to confirm his/her guesses, and then generates new predictions about what will come next (Goodman, 1988). Goodman (1967, p. 13) argues that efficient readers utilise knowledge of syntax and semantics to reduce their reliance on the graphic cues of the text.

Despite the outstanding contribution the top-down theories have offered to our understanding of how reading operates, several authors point out that top-down models also suffer from a number of limitations. Eskey (1988), for instance, warns that the top-down approach puts a strong emphasis on high-level skills, at the expense of lower-level skills and it claims that the reader samples only small parts of the text. However, evidence from eye movement research indicates that there is fixation on most of the words in a text (Rayner & Pollatsek, 1989). Another criticism of top-down theories is centred on the claim that good readers are more proficient than less-skilled readers in their guessing and reliance on context (Urquhart & Weir, 1998). This assumption is refuted by Stanovich (1980), who examined a large number of reading studies, and provided evidence that good readers use context less than poor and unskilled readers do. This can be attributed to the fact that compared with less proficient readers, who have very limited lexical knowledge, good readers know more words in the text (Juel, 1999).

The overemphasis that top-down processing puts on higher-level processing poses a further problem. Readers who have little knowledge about the topic of the text being read may find it difficult to generate predictions about the upcoming events (Eskey, 1988; Samuels & Kamil, 1988). Accordingly, Eskey (1988) points out that top-down models provide an accurate description of the processes of reading used by skilful, fluent readers who have developed their automatic perception and decoding skills. However, these models cannot account for beginner readers who may favour attendance to more bottom-up features.

It is clear from this analysis that both the bottom-up and top-down models appear to lack a detailed explanation of the reading process. This has consequently led others (Rumelhart, 1977; Stanovich, 1980) to formulate theories which combine the strengths of these two models and offer a more comprehensive description of reading process: *the interactive approach*.

2.3.2.3 Interactive processing in reading

According to interactive processing theories, reading operates between the bottom-up and the top-down models and it is supposed that certain perceptual and cognitive process take place during the reading process in either a parallel or an intermittent fashion. Here, what distinguishes the interactive model from the bottom-up and top-down models is that the former view holds that the reading process is an interaction between the textual information and the information the readers bring to the texts (Eskey, 1988). In this sense, the reader is not seen as passive, but has an active role to play in the reading process, as he/she is constantly making hypotheses about the text being read. Barnett (1989) states that interactive models are suggested to broaden our understanding of the relationship between the readers' vocabulary skills and comprehension. These models recognise that the comprehension of the text relies on the words in the page, despite the influential role these reading theories give to higher-level processing stages compared with lower-level stages.

Rumelhart (1977) strongly criticises bottom-up models (particularly the models by Gough, 1972; LaBerge & Samuels, 1974) which allow one *linear* direction of information processing, from lower-level to higher-level. In response to that, Rumelhart was the first to offer a purely interactive model of reading, stating that reading is "at once a 'perceptual' and 'cognitive' process" (1977, p. 573). The strong point of Rumelhart's (1977) model is the concept of related processes, which allow a parallel combination of higher-level and lower-level information processing (Samuels & Kamil, 1988; Tracey & Morrow, 2006).

That is, the reading process entails both sensory input (letter features) and nonsensory information about the orthographic structure of the language, which meet in one place and allow the reader to simultaneously use several knowledge sources in word recognition. The model proposes that *graphemic input* enters the process and is registered in a *visual information store* (VIS), “a sensory store that maintains visual information for approximately one-quarter of a second” (Reed, 2007, p. 32). This cognitive feature extraction device is then involved in the choice of the most important features from the graphemic input. The selected input at that point is brought to a *pattern* synthesiser along with syntactical knowledge, semantic knowledge, orthographic knowledge, and lexical knowledge, which work simultaneously together to facilitate the construction of the most probable interpretation of the words.

Based on the theoretical principle of background knowledge and the influential role it plays in reading process, Bamford (1984) developed a new schematic-theoretical model: *schema theory*. According to this theory, all of the knowledge an individual has acquired is stored in the memory as units or schemata. These knowledge units or schemata are described as *mental file folders*, in which information about particular events, ideas, or concepts is packaged (Honeycutt & Bryan, 2010). Rumelhart (1984) suggests that interpretation and perception are central to memory, therefore the basic notion of this theory is that the readers constantly evaluate hypotheses about the most probable interpretation of a text. Consequently, the new information in the text will be understood if it is integrated with existing information (schemata), offering a coherent account of the various aspects of the text. If readers find the new information confirmatory, they will continue and further elaborate their hypotheses. Otherwise, “they eliminate the hypothesis and construct another, which is consistent with the input data” (1984, p. 17).

Based on Rumelhart’s (1977) view of the reading process, Stanovich (1980) developed what he called the ‘interactive-compensatory model’ to explain

individual differences in reading. In this model, reading is described as a combination of processes in which the reader uses multiple sources of information to facilitate comprehension. Stanovitch's model stresses that not only do cognitive and information processing interact, but also that "a process at any level can compensate for deficiencies at any other level" (1980, p. 36). As noticed by Stanovitch, beginner readers may use their knowledge and the sentence context to compensate for a weakness in recognising the unknown words they encounter in a text. Thus, this model is interactive in its widest sense because its basic assumption is that sources from two dimensions (bottom-up and top-down processing) can combine in a hugely interactive fashion (Grabe, 1988). In a very real sense, on the other hand, Stanovitch's model is compensatory, as it suggests that well-developed sources can also compensate for less familiar ones when reading is difficult (Urquhart & Weir, 1998).

Like other types of reading model, interactive theories have also received criticism. As indicated Rayner and Pollatsek (1989) for instance, Rumelhart's (1977) model cannot be considered as a comprehensive model of reading due to its failure to account for the basis upon which different kinds of hypotheses are generated. Moreover, the model does not explain the extent to which different sources of knowledge are important in the interaction process. A further criticism is levelled at schema theory. Carver (1992, p. 165) mentions that many schema-theory based studies are mostly linked to hard reading materials. Therefore, he argues that schema theory is valid for reading relatively difficult materials, but it has little to say about "normal, typical, or ordinary reading".

To sum up, in comparison with the bottom-up and top-down models, interactive theories appear to constitute a widely accepted paradigm by both L1 (e.g., Kim & Goetz, 1994) and L2 researchers (e.g., Bernhardt, 1991), despite the aforementioned flaws. Reading research has shown that it is

difficult to draw a clear distinction between first and second language reading processes (Davies, 1995). However, it is suggested that no single theory can be said to be fully generalizable, and in certain conditions, some lower level decoding is more important and effective than higher-level processing and vice versa. As argued by Paran (1996), in order to compensate for their limited linguistic skills, the beginner L2 reader, who relies more on context, could find bottom-up processing more effective than L1 readers. Using their experience of the world and knowledge of the language, more advanced L2 learners may find a top-down model more appropriate for their reading.

Like many Arab EFL learners, Libyan students seem to encounter some difficulties when they read in English (the next section will shed light on these problems). Mourtaga (2006) found that Arab students tend to be poor readers in English because they have had no prior exposure to the English reading materials and consequently find reading materials something new and hard to deal with. “While they think that all of the meaning lies in the print, they do not have the linguistic competence to utilize that” (2006, p. 5). Consequently, Libyan university students, in my experience, appear to largely rely on the bottom-up reading approach which allows them to read the text word by word and sentence by sentence. Because Libyan university students are not involved in the selection of reading texts, they may find the materials very hard to read and understand, in addition to lacking interest in the subject. Therefore, one possible way to help the Libyan EFL students to use and develop a more interactive approach, involving both bottom-up and top-down strategies, is by allowing students to extensively read simple and interesting reading texts with easy vocabulary, which have been specifically chosen to suit their level. Graded readers seem to be an effective resource that can be used for such purpose as they provide L2 learners with simplified texts written to cater for their different levels of language proficiency.

Because the present study uses ER to measure incidental vocabulary gains, issues related to this reading approach and graded readers will be discussed in section 2.3.4. Prior to that, however, some common reading problems faced by Arab EFL students will be discussed.

2.3.3 Problems faced by Arab EFL learners when reading in English

Recently English has gained significant importance in Arab countries. The increasing popularity of this language is due to its status as an international medium of communication and its vast application in the business and corporate sectors. Still, the need for English differs from one Arab country to another. It seems that many EFL learners in some modern Arab societies (e.g., the United Arab Emirates) enjoy an advantage in language learning over many of their fellow learners in other Arab countries. This is due to the variety of opportunities they have to encounter or use English outside the classroom (O'Sullivan, 2004). Despite the fact that the importance of L2 language exposure outside the classroom in developing the learners' L2 competence is well recognised, Patil (2010) claims that reading remains the major source of language input in both ESL and EFL settings. English language learning in Arab countries is no exception, which may explain why reading dominates EFL learning and teaching materials in these countries. In these classrooms, teachers of reading spend a considerable amount of time on reading texts aloud to their students. While doing this, it is very common that the teacher explains the meaning of the text to their students by going through it sentence by sentence. Furthermore EFL teachers also very often ask students to perform additional reading at home (Patil, 2010). Nevertheless, research has repeatedly reported that Arab EFL learners face several problems while reading in English.

Reflecting on these problems and their causes, the discussion in this section will be based on four categories identified by Mourtaga (2006). Mourtaga

studied the reading skills of adults Arab students in the Gaza Strip and classified the problems that Arab learners of English face when reading and understanding EFL texts into four categories. Generally, the reading problems of Arab EFL learners are related to 1) the misunderstanding of the reading process, 2) insufficient competence in and use of English, 3) the differences between Arabic and English and 4) the English spelling/ sound system (Mourtaga, 2006, p. 1). The scope of this section will be restricted to discussion of these categories as they seem, from my experience, to cover most of the reading challenges facing Arab EFL learners.

2.3.3.1 Problems of misunderstanding of the reading process

According to definitions of the reading process, meaning is constructed when the reader interacts with the printed page (Anthony et al., 1993; Grabe & Stoller, 2001). Therefore, it is argued that misunderstanding of this process may result in major reading and comprehension difficulties. In most Arab countries, EFL teachers adopt the traditional bottom-up theory and the Grammar translation approach in teaching their students reading skills (O'Sullivan, 2004; Mourtaga, 2006; Al-Homoud, 2007). Therefore, many of these teachers consider reading as a one-way process, thus they put great emphasis on word identification. Such emphasis can be noticed particularly in traditional Arab EFL reading classrooms, where the teacher very frequently stops to correct or ask students about meaning of a particular word they have encountered while reading aloud (Mourtaga, 2006).

Arab EFL students are very often found to have limited experience with active reading and limited background knowledge about the subject (Al-Issa & Dahan, 2008). Furthermore, Patil (2010) mentions that many Arab EFL learners are addicted to the bad habit of sub-vocalisation. This sub-vocalization happens when the reader grasps the meaning of the texts s/he is reading by internally pronouncing (in the mind) the letters and words in the sentences. Although learners may not move their lips or speak words aloud, sub-

vocalization greatly slows down the reading process, as the mind is busy looking for meaning (Patil, 2010). Likewise, Patil, (2010) observes that eye fixation (actual stops or pauses between eye-spans when the eye is moving to its next fixation point) and regressions in reading can cause a significant reading speed problem for Arab EFL learners. According to him, Arab learners are slow readers because they fixate word by word. Likewise, because of the number of unfamiliar words in the texts, they tend to jump back to earlier words on the page in order to check comprehension (regressions). Research on eye fixation has shown that the reader's "saccade length can be as short as 1 letter space and as long as 15– 20 letter spaces" (Rayner, 2009, p. 1460), and evidence from L2 reading indicates that fluent readers look at phrases and groups of words (chunks) rather than individual words (Yamashita & Ichikawa, 2010).

Moreover, according to traditional methods of teaching reading in Arab EFL classrooms, learners are often required to read aloud in turn and, while reading, they use their index fingers or sometimes a pen to follow the words they read. Despite the evidence indicating that reading aloud improves comprehension of English native children (Amer, 1997), the negative impact of this teaching method on the Arab learners' reading abilities is strongly stressed by Smith (1971, p. 160), who argues that it slows reading speed down and interferes with understanding the meaning of the text being read. In the same way, Mourtaga (2006) believes that these methods of teaching EFL reading encourage the students to develop slow reading habits for the rest of their lives. The fact that the bottom-up approach has been adopted in many EFL classrooms in the Arab world indicates that many EFL learners do not have sufficient high-level processing (top-down) skills, which would allow them to involve perception and predictions in the reading process and therefore become faster readers (Oakhill & Beard, 1999).

2.3.3.2 Problems of insufficient competence in and use of English

Grabe (2009) emphasises that without gaining a reasonable amount of linguistic knowledge, learners will struggle to read and comprehend L2 texts. Saville-Troike (1984) argues that a reader's L2 lexical knowledge is a crucial factor in successful understanding of texts, and even more important than other types of linguistic knowledge, including syntax. This importance of vocabulary knowledge in reading comprehension has been stressed not only in L2 research (Cooper, 1984), but also in language research examining L1 reading (Freebody & Anderson, 1983). Arab students learning English as a foreign language seem to suffer from this problem. As Mourtaga (2006) points out, "[w]hen an Arab EFL student is asked why s/he finds English a complicated language to read, the common answer is that most of the words in the text are new and s/he does not know what they mean" (p. 5).

Moreover, Donnell and Wood (1999) suggest that L2 learners' difficulties with reading comprehension are generally related to aspects of the reader himself, the written page, and the readability of the text. In relation to the readability factor, these authors consider word difficulty and the length of sentences in the texts as two major problems that face the L2 reader, in terms of understanding the conveyed message in the text. Arab EFL learners are likely suffer from these problems, which can prevent them from becoming proficient readers and limit their competence in reading comprehension and reading speed. The inadequate proficiency of Arab EFL learners' in reading English has been confirmed in a number of studies conducted in several countries including Saudi Arabia, Sudan, Jordan, Yemen, the West Bank, Yemen, Syria (Fareh, 2010; Khan, 2012), Libya (Pathan & Al-Dersi, 2013), Algeria (Azeroual, 2013), and Egypt (Wahba, 1998, cited in Jdetawy, 2011). For instance, Azeroual (2013) studied the relationship between Arab students' knowledge of general English language and their comprehension of science texts. This study demonstrated that insufficient knowledge in English grammar and vocabulary

was the major obstacle that prevented adult Algerian EFL students from achieving a substantial level of reading comprehension.

A number of researchers (Al-Khasawneh, 2010; Fareh, 2010) ascribe the Arab students' limited EFL linguistic knowledge and competence to the poor teaching and the methodology adopted by many teachers in the Arab world. In addition, Mourtaga (2006) attributes it to the learners' insufficient EFL reading practice as well as their limited exposure to the English language.

2.3.3.3 Reading problems due to the differences between Arabic and English

Another problem, which many Arab learners seem to face when they read in English, is related to the differences between the Arabic and English languages. As stated by Patil (2010), the Arabic language is not one of the Romanised languages, so for Arab EFL students, dealing with the English alphabet appears to pose a real challenge to their reading and writing. The difference between the writing systems of English and Arabic languages affects the way Arab speaking learners read the script in English. The most notable distinction between the two languages is that English, like all Roman languages, is read from left to right; Arabic, in contrast, is one of very few languages written from right to left (Thompson-Panos & Thomas-Ruzic, 1983). Therefore, when reading, Arab EFL learners may find it difficult to reverse their habitual visualisation and subsequent regressions (Mourtaga, 2006).

Furthermore, unlike the English language, Arabic does not utilise a word capitalization system for its 28 letters. However, each Arabic letter has as many as four different shapes which are used differently according to the position of the letter in a word (Thompson-Panos & Thomas-Ruzic, 1983; Mourtaga, 2006).

Ryan and Meara (1991) add that the a significant difference between English and Arabic words lies in the fact that Arabic lexical vocabulary is very different from the lexis in Indo-European languages,

where words tend to be made up of a relatively stable root, and a system of affixes that are added on to this stem. Most Arabic words are based instead on a root that consists of three consonants, and these three consonants can be combined with different patterns of vowels to produce a whole family of words that share a common meaning (1991, p. 533).

When talking about the consonants in Arabic language, it should also be noted that differences between the two languages in this aspect also exist. There are a number of consonants that exist in English and not in Arabic and vice versa. This can undoubtedly cause a serious challenge for the learners of both languages. For example, because Arabic language does not have the letters [v] and [p], EFL learners are very likely to struggle with pronouncing words that contain these letters (Mourtaga, 2006). Similarly, the structure of consonant clusters in English seems to cause another reading problem for Arab EFL learners. Mourtaga (2006) refers to the fact that English language rules of pronunciation, allow up to three consonants to separate two vowels in one single syllable, but Arabic writing rules allow only up to two. Arab learners also encounter some difficulties in applying these rules to English words, so, in order to facilitate the English pronunciation of consonant clusters, they tend to insert an additional vowel between them. For example, Egyptian EFL students frequently “add an epenthetic vowel /i/ to avoid two- or three-constant clusters”. A good example is the word ‘children’ /tʃɪldrən/, which they pronounce as ‘children’ /tʃɪldirən/ (Cook, 1996, p. 43).

2.3.3.4 Problems due to the English spelling/ sound system

The dissimilarity between English and Arabic orthography causes a further problem for L1 Arabic speakers in learning English. El-Dakhs and Mitchell (2012) point out that the complicated nature of English spelling, which has no clear sound-to-grapheme rules, creates a major problem for Arab EFL students

developing their reading competence in English. It is generally known that Arabic is one of several languages which are orthographically sound based (Cook, 1996), whereas there is no clear correspondence between the English alphabet and its sounds, as the 26 letters can refer to about 43 sounds (Mourtaga, 2006). As noted by Fender (2008, p. 28), “L1 Arabic literacy skills develop in the fully-vowelized (i.e., transparent) script with reliable and consistent grapheme-phoneme mappings”. Therefore, as EFL language learners, Arab students may experience difficulties reading English words, which include certain complex correspondences between sound and letters such as *through*, *caught*, *schedule*, and *picnic* (Cook, 1996).

Pathan and Al-Dersi (2013) investigated the major reading problems encountered by Libyan university EFL students. The study showed that 80% of the subjects considered pronunciation as a major factor hindering their reading aloud in English. Pathan and Al-Dersi (2013, p. 2) identified various causes for this problem including that, a) several English sounds do not exist in the sound system of Arabic; b) the subjects’ experience of learning vocabulary is based on the way a word is spelt and not the pronunciation rules; c) they experience a lack of native English language exposure both inside and outside the classroom. In another study Wahaba (1998, cited in Jdetawy, 2011) also found some similar problems faced by Egyptian EFL learners at university level. According to Wahaba, the most noticeable pronunciation difficulties of Egyptian EFL students are those related to intonation and syllable stress. He continues to ascribe this to the difference in the pronunciation systems between Arabic and English, which can impact on the readers’ habit of sub-vocalisation and further slow their reading.

Taking the above mentioned problems into account, it is obvious that reading in English for many Arab EFL students is a nightmare. Mourtaga (2006) provides some suggestions for helping the Arab EFL students to overcome these reading problems. As stated by Mourtaga, EFL teachers need to introduce

their students to more effective teaching methods and techniques based on the Extensive Approach to reading. This can be implemented along with classroom instructions aimed at developing the skills and strategies required for proficient reading.

Hargis (1999) argues that readers of English, including Arab students, can improve their reading skills if their instructors adopt a more extensive and broader approach while teaching these students how to read. Day and Bamford (1998) suggest an extensive approach to help L2 learners to learn how to read, which fundamentally involves reading easy and interesting materials for pleasure, and reading for general meaning, rather than detailed comprehension. Similarly, Duffy (1997) argues that no reading method is superior to the other; rather, teachers need to adopt an approach that allows free and enjoyable reading, in order to teach students how to read properly in English. By adopting such an approach (extensive reading), teachers can ensure that the students are fully involved and with complete interest, therefore considering reading as fun rather than as a compulsory subject.

The next section will present the ER approach, with a specific focus on issues related to its underpinning theory, principles, graded materials, as well as the advantages and drawbacks of its implementation.

2.3.4 Extensive reading

2.3.4.1 Definitions

ER differs greatly from the more traditional intensive reading approach. When involved in intensive reading, learners typically read a short text and the teacher's role is to help and guide them to obtain the detailed meaning of the passage, to perform some vocabulary and grammar learning activities, and to develop their reading skills. ER, in contrast, refers to a situation where a language learner selects, and reads on his/her own, a large quantity of texts,

which are either at or slightly above their comprehension or linguistic level (Krashen, 1993; Day & Bamford, 1998; Nation, 2001). A very broad definition of ER is that it “generally involves rapid reading of large quantities of material or longer readings (e.g. whole books) for general understanding, with the focus generally on the meaning of what is being read than on the language” (Carrell & Carson, 1997, pp. 49-50). From a different viewpoint, Davis (1995) offers a more elaborated definition, that deals with the practical aspects of an ER classroom:

An extensive reading programme is a supplementary class library scheme, attached to an English course, in which pupils are given the time, encouragement, and materials to read pleurably, at their own level, as many books as they can, without the pressures of testing or marks. Thus, pupils are competing only against themselves, and it is up to the teacher to provide the motivation and monitoring to ensure that the maximum number of books is being read in the time available. The watchwords are quantity and variety, rather than quality, so that books are selected for their attractiveness and relevance to the pupils’ lives, rather than for literary merit (1995, p. 329).

The general picture that emerges from these definitions is that the most important features of ER are quantity, reading for information, general understanding, fluency and pleasure. It is the type of reading that students do in their leisure time and outside the classroom. The most important aspect of this type of reading is that learners select their own texts, which they believe they will understand and enjoy, on their own. Based on these features, a number of researchers have termed ER in different ways. Mason and Krashen (1997) for instance use the term ‘free reading’ or ‘free voluntary reading’ (Krashen, 1993), and Grabe (1991) calls it ‘sustained silent reading’, whereas Mikulecky (1990) terms it ‘pleasure reading’ and Elley and Mangubhai (1981) call it ‘book flood’. While these scholars have chosen different names for extensive reading, all terms given seem to have the same theoretical basis, which emphasises reading as pleasurable practice that should involve interesting, easy and largely self-selected texts.

2.3.4.2 Theoretical background

The underlying theory of ER is that language is learnt when the message is understood (The Input Hypothesis by Krashen, 1985). According to IH, learning is more likely to take place when we understand what we hear and when we comprehend what we read. More recently, Krashen (2004) has developed another theory, which he refers to as Comprehension Hypothesis. Comprehension Hypothesis suggests that three conditions are needed for acquisition to take place: a) there should be rich input available for learners (that means learners need to have a wide range of learning materials), b) this input should be comprehensible, and c) the input should be slightly above the learner's current language level. Although Krashen (2004) did not give enough explanation of how comprehension should be defined or what conditions are required to achieve such comprehension, in his earlier IH (Krashen, 1985, 1991) he suggests the $i+1$ concept, where i refers to the learner's current level of proficiency. Likewise, Krashen proposes that comprehensible input must be at a +1 level, which means that input should comprise materials from a level just above the assumed level of the learners. Taking reading as an example, the comprehension hypothesis suggests that in order to promote further acquisition of language, learners must read comprehensible texts at a level that is a little beyond their present acquired competence.

The IH states that the learner should focus on the meaning and not the form of the message. If the theoretical foundation of ER is considered, it is obvious that it is built on Krashen's comprehensible IH. The key principles of ER put the focus on input rather than output, and its success is based on the amount and the level of the input. ER involves the learner in a continuous reading process, where it is not necessary that the reader understands every word he/she reads, as the objective is to obtain an overall understanding of the message and the general context of the text (Bamford & Day, 1997). More importantly, the reading is individualised. In other words, learners choose the materials that interest them and read them autonomously without the support of the teacher,

as would be the case in intensive reading. This results in other features of ER, such as the fact that there are no follow-up tasks after the reading and learners are free to stop reading and choose other texts if they find the text is not interesting, or when the text they read is inappropriate to their level of competence (Bamford & Day, 2004). The next section highlights the most common characteristics of extensive reading.

2.3.4.3 Characteristics of extensive reading

ER is based on the well-established principle that we improve our general language proficiency in both first and second languages by reading. Several attempts have been made to provide some insight into this concept. The most widely recognised concept in the literature of ER is that introduced by Day and Bamford (2002). The authors present what they call the "ten top principles" of ER as follows:

- The reading material is easy.
- A variety of reading material on a wide range of topics must be available.
- Learners choose what they want to read.
- Learners read as much as possible.
- The purpose of reading is usually related to pleasure, information and general understanding.
- Reading has its own reward.
- Reading speed is usually faster rather than slower.
- Reading is individual and silent.
- Teachers orient and guide their students.
- The teacher is a role model of a reader (2002, pp. 137-140).

This comprehensive list offers a good guideline for conceptualizing ER in language teaching and learning. In response to this, Macalister (2010, p. 63) categorised Day and Bamford's principles into four sets. According to him, elements 1 and 2 in the list determine 'the nature of reading material', whereas

5 to 8 determine the reading *per se*. Points 3 and 4 consider the learners' behaviour towards the reading material and 9 and 10 deal with the teacher's role in the ER programme. As for understanding what ER is, it is felt necessary to briefly discuss these principles as they were originally ordered by Day and Bamford (2002).

1. The reading material is easy.

This is not an indication to encourage using material that is clearly below the learners' language level, but the text must be well within their reading competence. Nation (2001) suggests that in order to achieve the goal of suitable pleasurable extensive reading, the density of unknown words should not exceed one in every fifty words. Nation's proposal seems to be supported by what Krashen calls $i + 1$ in his IH (Krashen, 1985, 1991) (i.e. 'I' indicates the learner's current proficiency level and 1 means the language level slightly above the learner's current level). It is assumed that such a slight change in the degree of difficulty will serve to promote the guessing of the meaning of unknown words from the context and reduce the number of occasions when a dictionary might be used. Bell (1998, p. 3) warns that the constant looking up of new words in a dictionary will result in "slow, inefficient reading and destroy the pleasure that reading novels and other literature is intended to provide" (Bamford & Day, 1997, p. 3).

2. A variety of reading material on a wide range of topics must be available.

ER is based on the principle of reading for one's own pleasure, at an appropriate level for one's language proficiency. Therefore, in order for an ER programme to succeed, its library should make a variety of texts available for learners, in order that they can select what they really like to read. This should include a broad range of simplified printed materials written for L2 learners. Day and Bamford (2002) suggest that an ER programme should ideally comprise not only fiction and non-fiction texts, but also materials such as magazines and newspapers. Because of the lack of newspapers and magazines

written for L2 learning purposes, many ER programmes have typically used solely graded fiction and non-fiction books. Due to their variation in language level and genres, the graded books can inspire a flexible approach towards reading, since the reading is aimed at gaining general information and learners read for different purposes. Such variety in texts can also help learners to build awareness of other cultures, values, and beliefs, which, in turn, can increase their motivation and encourage them to read (Day & Bamford, 2002). In order to offer this condition in the current study, a collection of 40 different graded readers (by Oxford Bookworms) at different levels were offered to the ER group to select and read from. These titles were chosen from different genres to give the participants an opportunity to read topics of their interest. The benefits of using graded readers in this research are discussed in section 2.3.4.5 below.

3. Learners choose what they want to read.

ER is based on the notion of reading interesting texts for pleasure. Therefore, learners should be allowed to select from a wide range of textual materials that are interesting and suitable for their competence. They should also be free to choose the time at which they would like to read. If they feel tired, bored, or the text is either too difficult or too easy at any point of reading, then they should be encouraged to stop reading (Day & Bamford, 2002). In the present study, participants had the chance to select titles that interested them, but due to the research design, they had to adhere to reading their selected books during the programme. Because this study investigated pre-and-post reading incidental vocabulary growth, it was infeasible to allow participants to change levels or titles of books during the ER programme. Therefore, the principle of learners' freedom to read what they want was to some extent limited by the way this experiment was carried out.

4. Learners read as much as possible.

In order for ER to be an effective means for language learning, learners need to read as much as possible. This assumption is inspired by the fact that we learn

by reading (Nagy et al., 1985; Nagy, 1997). The more learners read of appropriate and interesting texts, the greater the benefits (Nation, 2001). Therefore, learners should be encouraged to read a large number of texts in their free time, so they can incidentally improve their language proficiency levels.

5. The purpose of reading is usually related to pleasure, information and general understanding.

Reading should be for its own sake, and learners must not feel that they are performing a learning task. For example, learners are very likely to suffer certain levels of stress when they are involved in intensive reading, if they expect to follow this with comprehension tests. In contrast, ER should be a stress-free approach and reading is mainly for pleasure and general understanding purposes. Hence, learners should not, under any circumstances, feel they are reading to satisfy a course requirement (Davis, 1995). Clarity (2007) considers it an important part of an ER programme to ensure that students are given access to texts of their interest. This can make the reading process enjoyable and motivate learners to sustain reading. In this study, ER was thought not to be familiar to Libyan EFL learners, so this point was addressed with care. Before they started the reading programme, the participants were provided with sufficient information about the ER approach. In order to ensure they would not be affected by their classroom practice of intensive reading, it was important to make them aware that the reader in ER reads for pleasure, information and general understanding, therefore reducing the stress-level they could have experienced in their traditional reading classroom.

6. Reading is its own reward.

ER aims at involving learners in experiencing reading that reflects reading in everyday life. That is, learners read in silence and at their own pace. The most important feature which differentiates ER from intensive reading is that learners are not required to answer comprehension questions or explicit

exercises or tests. The overall goal of reading is seen as pleasure and gaining a general understanding, so they should read without any pressure of testing afterward (Wilkinson, 2010).

7. Reading speed is usually faster rather than slower.

The self-select method of ER material allows learners to choose and read texts that are typically interesting and within their linguistic ability. Moreover, in extensive reading, learners are typically discouraged from using dictionaries, which interrupts reading and affects fluency. They are, however, advised to infer the meaning of few unfamiliar words, or possibly ignore them when they meet them in the text. In such cases, reading could become a smoother process which accordingly increases reading speed (Day & Bamford, 2002). Bamford and Day (1997) claim that ER provides learners with several advantages, including enhancement of reading speed. Nation (2005a) also suggests that graded readers are very effective for encouraging greater reading speed, as they are much easier than the level of texts that learners typically read to gain meaning-focused input in class.

8. Reading is individual and silent.

Individualisation is another distinguishing feature of extensive reading. The most important aspect is that students are allowed to individually select interesting materials from within a variety of topics. Similarly, they need to read these self-selected texts individually and in silence. Although there is an argument emphasising the importance of oral reading in improving learners pronunciation skills (Nuttall, 1996), silent reading is seen as a more beneficial means for improving the skills of reading comprehension, reading speed, and word meaning guessing (Bamford & Day, 1997). Another benefit of silent reading is that it allows learners to explore the notion that reading is a personal interaction with the text (Day & Bamford, 2002), which can increase their confidence in their ability to independently read and understand texts (Hopkins, 1997).

9. Teachers orient and guide their students.

A successful ER programme entails some initial work from the teacher. It is important for the teacher to ensure that students are given an early introduction to extensive reading, particularly in classrooms where ER is an absent approach. Several issues regarding reading materials, the independent nature of text selection and reading should be discussed with learners before commencing an ER programme. Teachers also need to raise the learners' awareness that this can lead to improvement in reading skills, as well as overall language learning proficiency, so they will be aware of why they should read extensively.

10. The teacher is a role model of a reader.

The teacher needs to act as an active reader and read the same materials in order to stimulate students' reading. This also helps the teacher to demonstrate to their students what it means to be a reader, and the advantages that they can gain from practising reading. Maley (1999, p. 7) emphasises the influence teachers have on their students, stating that "students do not just (or even) learn the subject matter we teach them; they learn their teachers". In his words, Nuttall (1996, p. 229) reminds teachers that "reading is caught, not taught". According to Prowse (2002, p. 144), the teacher can share their reading experience through group or individual discussion with their students. This kind of discussion can also provide "a natural ground where communication in English can take place easily and naturally".

Generally, these principles appear to highlight most of the aspects and satisfy the goal of extensive reading. Thus, it seems crucially important for teachers to take them into account when setting up an ER programme. In the present study, the ER programme was carried out after careful consideration of the issues raised above, but, for research purposes, principle 3 (see above) was adopted with certain care. More specifically, the participants selected what they wanted to read, but they were required to adhere to reading their selected titles.

Similarly, the principles (i.e., items, 9 and 10) concerned with the teacher's role (Macalister, 2010) were dismissed as irrelevant to this study, so they were not considered when the ER programme was carried out.

2.3.4.4 Some issues with setting up an ER programme

When setting up an ER programme, the first step, as Hill (1997) recommends, is to 'think big and start small'. This means that teachers considering setting up an ER programme should have a clear vision of what the programme might look like after several years. Waring (2011) indicates that many ER programmes end in failure because of the lack of clear direction and goals. Therefore, the very first step of any programme should be to seek to determine the most appropriate methods that will fit within the schools' overall goals and objectives. Waring also points out that:

If a successful program is to prosper, it has to have vision, and the will to survive potential threats to its existence. Among these threats are increases in lost or mislaid materials, insufficient resources to maintain a library, teaching and financial resources being moved to other projects, and a general lessening in enthusiasm after the highs of the "big start" (2011, p. 73).

Therefore, it is important that teachers establish clear objectives for their ER programme. Before planning an ER course, they need to find a way to make it fit with their school syllabus. In this regard, Day (2003) suggests three possible ways of introducing an ER course into an EFL curriculum. The first possibility is that ER can be introduced as a separate course. This can be implemented by integrating an ER programme independently from the school EFL curriculum, but allowing students to read both in and outside the classroom. The second possible way is through providing ER as part of an existing course. Adopting such mode will require the teacher and institution to include a certain amount of ER into a pre-existing course in the syllabus. The third possible method is by providing an ER programme as an after-school activity, in which students can be involved in an extra-curricular reading club. The teacher who runs the course may invite all students to take part in this reading club. In cases of

materials shortage, students might be asked to pay club subscription fees to contribute to the purchase of the reading materials. Day (2003) suggests that teachers can help students to create and maintain motivation and interest in ER by involving them in regular reading activities and raising their awareness of the reading goals. For instance, the teacher may hold regular meetings with students to discuss and reflect upon their individual readings. Moreover, the teacher may “organise regular reading marathons in which each student aims to read a certain number of pages during a set period of time” (Day, 2003, p. 2).

2.3.4.5 Graded materials for extensive reading

In L2 contexts, ER is largely limited to graded readers (Bamford, 1984; Hafiz & Tudor, 1990; Asraf & Ahmed, 2003; Hill, 2008). Although there is no reason to justify why such a connection exists, there are several benefits to using these simplified books as a major source for ER materials. First and foremost, in order for comprehension to take place, the reader needs to know an average of between 95% and 98% of the running words in the text (Hu & Nation, 2000). Similarly, Laufer (1989) found that comprehension was unsatisfactory when less than 95% of the text’s word tokens were unfamiliar to her subjects. It is obvious that such high percentage of word coverage is hardly achievable if L2 learners, whose language competence is below a certain level, are pushed to read authentic texts. Therefore, the only way to attain this level of comprehension is through simplified texts. Nation (2005b) concludes that, if the conditions necessary for learning from ER at the learners’ competence levels are to be met, it is important to use simplified texts. Graded readers appear to be the best means for ER purposes. A common rationale for using these materials is that they are written for the purpose of catering to different levels of language, so learners can select their level and read with ease. Texts are ‘graded’ in the sense that vocabulary and grammar structures are controlled to help the learners read with relatively fluency, without being overwhelmed by new words and structures (Hill, 2008). If the reading text is too difficult for readers, they will keep looking up many unknown words in a dictionary. This,

of course, will affect the fluency of reading, which is one of the key aims of extensive reading. The most popular publishers of English graded readers (i.e. Oxford Bookworms, Cambridge English readers, Penguin Readers, Macmillan Guided Readers) produce various titles of simplified books at different levels of difficulty to help L2 learners choose and read at their level of proficiency. These books are typically in a series with 4 to 6 different reading levels, graded from beginner to advanced (Bamford (1984).

There are other advantages of graded readers. Powell (2005) points out that the graded readers offer access to a wide collection of texts written in various genres. In other words, graded readers contain many kinds of simplified versions of various genres, such as original English novels and learners can select their own topics of interest. This kind of story, in certain instances, can help to develop learners' intercultural awareness by providing insight into other cultures. More importantly, unlike many EFL textbooks, which provide language in isolated chunks, graded books present language in context (Powell, 2005). Research shows that in comparison with learning from isolated language activities, context offers more meaningful input, which can help learners to learn how to form structures correctly and how to use them to communicate meaning (Nunan, 1998; Anderson, 2005). In this respect, it has been argued that more efficient processing becomes possible when a language learner deals with "related units of information rather than isolated bits" (McLaughlin et al., 1983, p. 138). At all costs, therefore, graded readers should be involved in the establishment of any ER programme. In the present study, a collection of graded readers from Oxford Bookworms will be used for the ER programme (see methodology chapter). The next section sheds light on some of the major features of ER which have led to this study being undertaken.

2.3.4.6 Benefits of extensive reading

Extensive reading as an educational approach has been broadly discussed over the past decades, and its benefits have been widely approved in a large body of

both L1 and L2 research. Huckin and Coady (1999, p. 182) describe ER as a “pedagogically efficient” approach to L2 learning as language learning occurs while the learners are involved in two activities: reading and acquisition. There are a number of reasons why it is so important to improve language knowledge through ER. Nation (1997) cites three simple reasons that make ER an attractive approach for second language learning. First, as already mentioned, reading is an individual practice through which learners of different language levels can read and learn at their own level and independently of their teacher. Second, learners do not have particular reading material imposed on them, but they are allowed to select what is interesting to them, which consequently increases their learning motivation. Finally, ER offers a great opportunity for language learning to occur outside the classroom. This specifically, is a significant advantage for students in EFL contexts where the target language input outside the classroom is characteristically very limited, as in Libya.

Numerous studies have shown convincing evidence of the substantial benefits of ER in developing the language knowledge of L2 learners. For instance, empirical research has demonstrated that ER has a positive impact on second language reading (Robb & Susser, 1989; Lai, 1993; Cho & Krashen, 1994; Masuhara et al., 1996; Kusanagi, 2004; Taguchi et al., 2004; Iwahori, 2008; Al-Homoud & Schmitt, 2009). Findings from these studies suggest that, like first language learners, second language learners can also ‘learn to read by reading’ (Nuttall, 1996, p. 168). Al-Homoud and Schmitt (2009) investigated the reading speed improvement of two groups of Saudi EFL learners’ in two learning settings: an extensive vs. intensive reading environment. Although the two groups improved significantly in their reading speed, the ER subjects reported a larger improvement rate than did the group in the intensive reading environment. Lai (1993) also examined the reading proficiency improvement of some Hong Kong EFL learners in secondary schools. The study indicated that ER helped these learners to significantly increase their reading proficiency.

Along similar lines, other researchers have studied the influence of ER on L2 learners' writing (Janopoulos, 1986; Hafiz & Tudor, 1990; Tsang, 1996; Mason & Krashen, 1997). Hafiz and Tudor (1990) asked English language learners in Pakistan to extensively read graded readers for a maximum of 90 hours. When compared with the control group, the ER subjects displayed significant improvement in their writing ability, including writing longer essays, improving their word spelling knowledge, using a variety of words in their essays, and using more accurate sentence structure. ER has also been found to improve motivation to read and the development of a positive attitude to reading in a second language (Elley, 1991; Cho & Krashen, 1994; Asraf & Ahmed, 2003; Nishino, 2007). Nishino (2007) conducted a longitudinal study on two Japanese secondary-school students to investigate the impact of ER on learners' motivation and learning strategies. Nishino found that her subjects had become highly motivated to read after two and half years of extensive reading. In a similar vein, Asraf and Ahmed (2003) involved students from three rural secondary schools in Malaysia in ER programmes and they concluded that ER leads to the creation of positive attitudes with regards to reading.

More significantly, incidental vocabulary growth has been widely recognised as the most substantial benefit of ER. This is perhaps due to the robust evidence gained from first language research (e.g. Nagy & Herman, 1985; Nagy et al., 1985; Nagy et al., 1987). More recently, ER has gradually gained popularity as one of the most pedagogically effective ways to promote L2 vocabulary development (Huckin & Coady, 1999). The value of this approach in learning vocabulary has been confirmed by research in worldwide contexts (Waring & Takaki, 2003; Horst, 2005; Pigada & Schmitt, 2006; Al-Homoud & Schmitt, 2009). For the purpose of this study, these studies and others involving short text reading will be reviewed in detail later, in Chapter 3. Next, some limitations of the ER approach will be highlighted.

2.3.4.7 Drawbacks of extensive reading

There are several weaknesses of ER as a learning approach. In terms of its potential benefit to incidental vocabulary gains, this assumption has been open to some criticism of its theoretical and practical aspects. Based on a review of early reading research, Raptis (1997) raises the question of whether incidental learning from reading is a viable approach since the available evidence points to a small contribution to L2 acquisition. Another argument concerning incidental word learning from ER is based on research on contextual inference strategies. Raptis (1997) indicates that it is very likely that L2 learners may make incorrect inferences because of unclear contextual information, and, as a result, they face the risk of acquiring words wrongly (Bensoussan & Laufer, 1984; Mondria & Wit-De Boer, 1991). A similar criticism is expressed by Huckin and Coady (1999), who argue that reading extensively for general meaning does not automatically result in vocabulary learning. In other words, an L2 learner might be able to make a correct guess of a word's meaning with the use of available contextual information, but this does not necessarily mean that the learner retains that meaning.

Davis (1995) also highlights some of the limitations associated with ER, but he considers the implementation aspect of the approach. One of the major problematic aspects of an ER programme, as highlighted by Davis (1995), is its cost. The preparation of an ER programme requires a large budget for purchasing books and other relevant materials. Unfortunately, many teachers and institutions are not able to cover such a high cost, which probably makes ER an underused approach in many ESL and EFL classrooms. Moreover, implementing an ER programme is not straightforward. In order to run a successful programme, there needs to be mass organization and paperwork. Teachers also need to adjust the reading that is done within the curriculum time. However, many teachers do not perceive such voluntary reading as a teaching method. Likewise, both teachers and students involved in teaching

and learning on a full-time syllabus are unlikely to find the extra time to spend on reading texts of their own interest (Davis, 1995).

Davis (1995) recommends that, in order to overcome these limitations, it is necessary for both teachers and institutions to first become convinced of the effectiveness of the ER in L2 teaching and learning. Even without a large number of graded books and other ER texts, it is good idea for the teacher to implement an ER programme, tailored to their students' needs.

From the above review of literature on extensive reading, it is obvious that there is widespread consensus in terms of the role of ER in promoting L2 language skills development, but there still seem to be doubts as to its effectiveness in relation to vocabulary improvement (Raptis, 1997). The small amount of vocabulary gain reported previously is perhaps due to the problems with the way in which most of the previous ER research has been carried out (see Chapter 3 for discussion of these problems). This study was conducted with great care, considering all issues of successful ER, in order to paint a clearer picture of the amount of incidental learning of vocabulary that occurs from ER.

Chapter 3:

Incidental vocabulary research

Chapter 3: Incidental vocabulary research

3.1 Introduction

This chapter reviews previous research on incidental vocabulary acquisition from reading. It aims at providing background knowledge of what has been done and discovered within the field. The discussion in this chapter will begin with a summary of reading studies conducted in the field of L1 acquisition, to offer some background on how the notion of incidental vocabulary learning was initiated. This will be followed by a review of the most relevant studies in L2 reading research. Studies looking at incidental vocabulary learning from reading short texts will first be reviewed, before studies using ER materials are discussed in detail, with a particular focus on the limitations. It is hoped that the knowledge drawn from this research will not only facilitate the process of conducting the present work, but will also expose some of the differences between the previous research and the present study, thereby enhancing the significance of this study.

3.2 Studies on L1 (First Language) acquisition

Since the early 1980s, several studies of L1 acquisition have focused on vocabulary acquisition from reading. It has been argued that much of English native speakers' vocabulary is acquired through incidental learning from context (Saragi et al., 1978; Jenkins et al., 1984; Nagy et al., 1985; Nagy et al., 1987). The assumption is that incidental vocabulary learning occurs at a very early age, while a child is engaged in verbal conversation. Only in the later stages, do they start to learn from reading activities. The most striking evidence for incidental learning from reading comes from the research of Nagy and Anderson (1984) and Nagy et al. (1985). These scholars found that during school years, children are often involved in free reading, therefore exposure through context makes a major contribution to vocabulary acquisition. This

vocabulary growth, however, seems largely determined by the child's experience with written language and their level of reading comprehension.

Based on a number of studies conducted with schoolchildren, a number of scholars (Nagy & Anderson, 1984; Nagy et al., 1985; Nagy et al., 1987) have estimated that in the course of a year, a child in the middle grades encounters approximately one million words in printed materials, every 1000 of which will include between 15 and 55 words unknown to the child. From calculations based on these estimates, Nagy and Herman (1987) conclude that a child encounters about 30,000 new words in a year.

As a consequence, there is a consensus that most L1 vocabulary is acquired incidentally from language exposure. Research has shown that native children initially learn new words from verbal communication with parents, but, at later stages, written exposure contributes to the further growth of learners' lexical knowledge (Jenkins et al., 1984; Nagy et al., 1985; Nagy et al., 1987). A number of researchers (Nagy & Anderson, 1984; Goulden et al., 1990; Nagy, 1997) have estimated that L1 learners acquire about 1000 new words every year. These findings suggest that, by the time L1 learners enter university, they usually know about 20,000 word families (Goulden et al., 1990). A higher estimate is given by Nagy and Anderson (1984), who propose that, during the school year, children between grades three and twelve learn as many as 3000 words, and that by the end of high school, a student will have acquired about 40,000 words. Similarly, White et al. (1990) investigated three groups of schoolchildren and concluded that their vocabulary knowledge increased by about 3000 words per year. Taking these figures into account, it has been argued that although the differences in these estimates may be due to the different methods used for classifying word units and families, only a small percentage of words are actually learned through direct teaching (Jenkins & Dixon, 1983). Instead, it is thought that a large proportion of this vocabulary is accounted for in terms of incidental acquisition from reading (Day et al., 1991).

From a review of the related literature, it appears that many studies have questioned the effectiveness of incidental acquisition from reading, when compared to other modes of learning (Jenkins et al., 1978). To investigate this widely debated question, Nagy et al. (1985) conducted an experimental study of eighth-graders, using natural texts, in which, of the 30 target words, 23 appeared only once. The research was of interest, as the methodology was specifically designed to measure small gains in vocabulary growth. When analysing the results in absolute terms, the researchers found that the contextual effect was small; however, the results were statistically robust and showed consistency across the texts, measurement methods and the level of scoring. As a result, it is suggested that a single, incidental encounter with a word is unlikely to result in complete understanding of its meaning.

Carey (1978) argues that if any vocabulary learning from context does occur, it must be perceived as incidental. This suggestion further implies that vocabulary learning is gradual in nature. Nagy et al. (1985) explain that, when this gradual process is further enhanced by exposure to written language, incidental acquisition of first language vocabulary can be significant.

Taking these issues into consideration, it can clearly be seen that incidental learning from context, and particularly from reading, plays a major role in first language acquisition. The L1 studies by Nagy and his colleagues established a theoretical assumption, which is widely acknowledged by L1 and L2 acquisition researchers. Their theory is based on the notion that incidental acquisition of words is incremental in nature, so word knowledge improves with repeated exposure through listening or reading activities. Huckin and Coady (1999), however, argue that incidental vocabulary learning predominantly occurs by means of reading. Such conclusions make this an important phenomenon, which has gained growing attention in the field of second language acquisition research.

3.3 Studies on L2 (Second Language) acquisition

The evidence from L1 research into incidental vocabulary growth from reading has encouraged many researchers (particularly in the 1980s and 1990s) to attempt to apply similar approaches to the field of second or foreign language learning. In this section, I shall be describing a number of influential studies. I shall go on to consider these studies critically, particularly with respect to their limitations and methodological flaws.

Early research in this area was concerned with incidental vocabulary learning from reading activities. Learners were typically given a short text to read, before being post-tested on vocabulary gains immediately after the reading (e.g., Pitts et al., 1989; Dupuy & Krashen, 1993; Horst et al., 1998; Rott, 1999a). More recently, however, studies have focused on incidental word learning through ER (e.g., Horst, 2005; Pigada & Schmitt, 2006; Al-Homoud, 2007; Al-Homoud & Schmitt, 2009). In such research, learners are exposed to a variety of texts at different levels, that they themselves have selected to read, and that suit their language abilities and interests. Many of the studies in these two strands of research (i.e. short text and ER research) have shown considerable incidental vocabulary gains. However, these studies have been widely criticised, due to the identification of several problems associated with their methodologies. Studies within these two research strands will be discussed in this section. The discussion will include some of limitations from which the studies suffered.

3.3.1 Incidental vocabulary learning from reading short texts

In this type of research, subjects are normally assigned to groups and given a text to read, before incidental acquisition of target words is measured (Horst, 2005). One of the earliest studies that looked at incidental L2 vocabulary learning from reading short texts was conducted by Saragi et al. (1978). In this study, 20 adult native English subjects read the novel 'A Clockwork Orange',

which comprised about 60,000 words in total, including 241 Russian slang words (*nadsat*). In order to measure incidental word learning from reading in this study, Saragi et al. (1978) tested 90 of the *nadsats*. Incidental gains in this study were measured through a multiple-choice test. After the reading, the subjects were asked to select the appropriate meaning for each of the *nadsat* words from choices of four English words. The results obtained in this study were highly encouraging, as the average score for learning of the 90 *nadsats* from reading the novel was 76%, with a 77% median. However, as Horst et al. (1998) caution, the case of native English subjects learning Russian words from an L1 context does not precisely represent the L2 learning condition.

This fact was of interest to other researchers. For instance, Pitts et al. (1989) replicated Saragi et al.'s study by asking two experimental groups of ESL learners (containing 35 and 16 subjects respectively) to read the first two chapters (about 6700 words) of 'A Clockwork Orange' for about an hour. After a 10-minute interval, the researchers tested their subjects' knowledge of 30 *nadsats* appearing in the chapters. Two control groups, of equal numbers (23 subjects), took the multiple-choice post-test with the experimental groups. Pitts et al.'s (1989) findings indicate that adult ESL learners were able to show a small but statistically significant incidental gain in vocabulary knowledge, when compared with the control groups. The first experimental group scored 6.4% (M=1.8, SD= 28) on the post test, while the second experimental group scored 8.1% (M=2.42, SD= 30).

Pitts et al. (1989) attribute the low incidental growth attained by their subjects to the difficulty of the text, since over 50% of them did not finish the assigned reading. Moreover, by assigning only two chapters of A Clockwork Orange to read, the researchers reduced the number of target item encountered, as the *nadsat* repetitions in the remaining chapters were excluded. In addition, not all of the *nadsats* appearing in the novel were tested, as other untested *nadsats* could have been acquired by the subjects in this study, so results could have

been underestimated (Pitts et al., 1989). The researchers further admitted, that the novel might not have held the subjects' interest, as it was not self-selected material.

Evidently, L1 learners learning *nadsat* words from a text written in their L1 language gave them an advantage over L2 learners who encountered the *nadsats* in L2 contexts. It is much easier for adult native speakers to use the surrounding text to guess the meaning of unknown words encountered when it is written in their native language, than for L2 learners to learn words from a foreign text. One conclusion to be drawn from the disparity is that level of comprehension is an essential variable in learning words incidentally from context. It is possible that, because more contextual clues were available to the subjects in Saragi et al.'s study, grasping the meaning of unfamiliar words incidentally was easier for them. In contrast, Pitts et al.'s subjects could have encountered severe problems with guessing meaning, when trying to access the meaning of unknown *nadsat* words in texts that may have included large amounts of new vocabulary.

Other studies have found evidence of incidental acquisition of L2 vocabulary by using a range of short texts, but all reporting similarly low rates of gain. Day et al. (1991), for example, carried out a study of two groups of 191 high school and 397 EFL students from Japanese universities. Students at each of the two educational levels were randomly allocated to a treatment and a control group. The unknown target words in the story were identified through a pilot vocabulary test, which was given to a group of participants who were similar to the research subjects and, who had read the same text. A word was excluded from the post-test if it had obtained more than 40% correct responses. The subjects in the two treatment groups read, in 30 minutes or less, an abridged version of a story called "Mystery of the African Mask" (1,032 words). Immediately after the treatment subjects had finished reading, they were given a multiple-choice post-test consisting of 17 words. The test was also given to

the two control groups, who had not read the story. The subjects were not permitted to use dictionaries and were asked not to guess the meaning of unknown target words. Instead, they were encouraged to select “I don’t know” when applicable. Neither the treatment, nor the control groups were informed that they would have a vocabulary test. The findings from Day et al.’s (1991) study showed that subjects in the treatment groups knew significantly more words than those in the control groups. The difference between the mean scores of the university treatment group (M= 9.34, SD=3.41) and the control subjects (M= 6.28, SD=2.97) was statistically highly significant, $p<0.01$. Similarly, the high school subjects in the treatment group scored significantly higher marks (M= 5.02, SD=3.21) than the subjects in the control group (M= 4.01, SD=2.48), $p<0.01$.

In their study, Dupuy and Krashen (1993) reported a much higher rate of incidental vocabulary gain. Their study investigated a group of 42 undergraduate students studying French at the University of Southern California. The researchers compared the performance of one experimental group with that of two control groups. The subjects in the experimental group were initially asked to watch the first five scenes of a play called ‘Trois hommes et un couffin’ and then, the next day, were given a forty-minute task that consisted of reading the script of the next five scenes (15 pages). For the purposes of the research, the film was shown without subtitles, to prevent the subjects from paying attention to any unknown lexical items. They were told that they would take a comprehension test after they had finished reading. The multiple-choice vocabulary test was designed to fit the number of word encounters in the text, and thus included 30 items encountered between one and six times. The selection was made by a native French speaker, on the assumption that the words chosen would be colloquial and therefore unlikely to be known by the subjects. The subjects, in all groups, completed the vocabulary post-test immediately after the experimental group had finished reading. Compared with the control groups, who had neither read the text, nor watched the film, Dupuy and Krashen (1993) found that the average gain of the

experimental group was nearly 6 words, which suggests that about 20% of the new words were learned in just 40 minutes of reading.

Horst et al. (1998) used a simplified novel ('The Mayor of Casterbridge') to measure incidental vocabulary acquisition by 34 lower intermediate EFL students at Sultan Qaboos University in Oman. The reading treatment lasted for ten days. In order to make sure that all subjects had read the whole story, Horst et al. (1998) implemented an unconventional reading strategy. The subjects had to follow the text in their books (21, 232 words in total), while their teacher read the story aloud. It was read at a pace that did not offer the subject any opportunity to look up unfamiliar words in a dictionary. In addition, the books were given to the subjects at the beginning of each session and collected afterwards. The aim of this technique was to reduce the possibility of new words being looked up in a dictionary or studied at home.

The target words were assigned from among a number of low-frequency English words identified by a computer program. The researchers assumed that, unlike high-frequency words, which were more likely to be encountered in the subjects' course books, low-frequency words were unlikely to be met anywhere but in the experimental text. Words occurring once in the text were excluded, due to being too infrequent to be good candidates for measurement of incidental learning over a short period. The list of target words, therefore, included eight items occurring more than seven times, and 37 items that appeared from two to six times in the text.

Horst et al. (1998) designed two test formats (multiple-choice and word association) which together were administered twice: as a pre-test a week before the reading and as a post-test immediately afterwards. The multiple-choice test consisted of all of the 45 items identified by the researchers, whereas the word-association test contained only 13 items. The results from the multiple-choice test indicated that the average "pick-up" of new words was

about one in every five tested words. Moreover, the subjects' performance in the word association test showed that they could make about 5 correct associations out of 13 items. A comparison between the subjects' performance on the pre- and post-tests showed a significant improvement in their knowledge of the vocabulary. The mean score in the pre multiple-choice test was 21.64, and jumped to 26.26 in the post-test, $p < 0.05$, while the mean in the word association test increased from 5.53 to 6.71, $p < 0.05$.

Clearly, the issue of how many exposures to a word are necessary before it can be considered as learnt needs to be further explored. In the context of learning a foreign language other than English, Rott (1999) conducted a study involving 67 German language learners studying at a university in the USA. In this study, the researcher used six treatment groups to explore the effect of exposure frequency (two, four, and six times) on the incidental learning of 12 unfamiliar target words. Rott divided the 12 target words into two sets, with three nouns and three verbs in each set. To prepare the reading materials, the researcher wrote six different short paragraphs of 4 to 6 sentences for each of the target words (72 paragraphs in total). The subjects were assigned to two treatment groups, and each group was exposed to one set of the target words. In like manner, every treatment group was further divided into three sub-groups. In order to examine the effect of exposure frequency on acquisition of the target words, the first subgroup was asked to read the paragraph containing the target words twice; the second subgroup read the paragraph containing the same target words four times, and the third subgroup encountered the word in six different paragraphs.

Before reading the paragraphs, Rott (1999a) made sure that all the target words were unfamiliar to her subjects. She administered a pre-test, using a checklist format, which contained 50 items: the 12 target words along with 38 distracters. The subjects were asked to provide translations of words they knew and skip those which were new to them. In addition, to assess the acquisition

and retention of the target words after reading, Rott used two post-test formats, which examined the receptive and productive levels of learning. For the assessment of word production, a supply-definition task was employed. The subjects had a list of the 12 target words and eight distracters, and were required to provide an equivalent L1 word meaning for these words. The subjects' responses were given a score of 0 (for unknown words), 1 point (for partially known words), and 2 points (for fully known words).

For measuring word acquisition and retention, Rott used a five-item multiple choice test, in which subjects selected from four possible definitions or an "don't know" option, when the meaning was unknown. For scoring this test, Rott (1999) gave 2 points to the correct choice and nothing was given to incorrect answers. The results obtained from this study indicated that two exposures could lead to significant acquisition of the target words. However, six exposures to the target words led to a greater number of significant gains than two or four encounters. Moreover, tests which assessed lexical receptive knowledge (multiple-choice and checklist tests) revealed higher retention scores than tests which measured word production (supply-definition test).

- *Critique of described research studies*

Taken together, the L2 studies discussed so far demonstrate that incidental acquisition of L2 vocabulary from reading is possible. However, apart from the study by Dupuy and Krashen (1993), the pickup rate reported in this research is relatively small. These early studies, however, seem to have been less effectively constructed than the studies with L1 subjects (Nation, 2001). Therefore, the research has been criticised for a number of limitations associated with the research methodology, meaning that the findings should be taken with caution.

The first problem concerns weakly designed tests. Despite the interesting findings, there remains a question about the instruments of measurement employed in these studies. Pellicer-Sanchez and Schmitt (2010) caution that some L2 reading research, which ignored the fact that vocabulary learning is incremental in nature, used vocabulary tests insufficiently sensitive to measure the learners' slight and partial learning of new words. In fact, all of the reviewed studies used multiple-choice tests as a technique for gains. However, Waring and Takaki (2003) warn that, for three main reasons, the multiple-choice test is not the most appropriate instrument for assessing the learning of new vocabulary. First, multiple-choice results can be affected by participants' random guessing. Second, this type of test only assesses "prompted meaning recognition" and is not effective in measuring "unprompted meaning recognition", which is most often needed for normal reading. Finally, it is always difficult for researchers to design a reliable multiple-choice test. A typical example of the drawbacks of MC tests, as argued by Waring and Takaki, is found in Dupuy and Krashen's (1993) testing instrument. It included four choices for each item: three possible answers and a "don't know" option. Although the "don't know" option could have served to reduce the chance of guessing, the poor question design in Dupuy and Krashen's study made this option ineffective. In their study, each question contained two semantically related words plus an odd one. Such a multiple-choice design is susceptible to guesswork, and can provide a 50% possibility of guessing the correct answer if subjects choose one of the two words of similar meaning and ignore both the odd word and the "don't know" choice. Likewise, the word association test in Horst et al.'s (1998) study presented three words, from which the subjects were asked to circle the odd one. It would have been better if additional distracting words had been given and subjects had been asked to identify any two similar ones. This, too, could have reduced the possibility of guessing.

Another problem with the testing was that some researchers (e.g., Day et al., 1991; Dupuy & Krashen, 1993) failed to determine the subjects' pre-existing target word knowledge, so as to establish the actual amount of newly-learned

vocabulary. Since the subjects' knowledge of target words was not pre-tested, it is possible that they knew some, if not all, of these words before the reading treatments began. This leads to a further, related problem with most of these reading studies. The incidental learning of words was measured immediately after the subjects had undergone the treatment (e.g., Pitts et al., 1989; Dupuy & Krashen, 1993). Since there was no interval between the reading and the post-tests, it is possible that at least some words were still in the subjects' short-term memory when they took the post-test. Schmitt (2010) cautions that, because vocabulary learning reported in immediate post-tests cannot necessarily be interpreted as long-term acquisition, the results of these studies must be assessed with care. A clearer picture of the actual gain could have been formed if other, delayed tests had been given to the subjects. Waring and Takaki (2003) reported that only a small number of newly learned words were remembered by the learners, when measured later by a delayed post-test.

The second major problem is concerned with either the amount of reading, or the way it was practised in the research discussed above. In general, the studies offered relatively short texts, asking the subjects to read over a short period of time. In some cases, the subjects read no more than six short paragraphs (Rott, 1999a), fifteen pages (Dupuy & Krashen, 1993), one excerpt (Pitts et al., 1989), or, at the very best, one short story (e.g., Saragi et al., 1978; Day et al., 1991; Horst et al., 1998). Accordingly, in most cases, reading tasks lasted between less than 30 minutes, (Dupuy & Krashen, 1993) and one hour (Pitts et al., 1989). Since repetition is an essential factor in promoting incidental word learning (Nagy & Herman, 1985; Nagy et al., 1985), the short readings provided in the earlier studies seem deficient in this respect. In their investigation of how words are learnt from reading, Nation and Wang (1999) suggest that, to ensure that new words are encountered again within a short period of time, a language learner should read one graded reader each week. However, this method is only possible if incidental vocabulary learning research involves extensive reading, as is the case in the current study.

The absence of longer reading tasks in many earlier studies led to the further problem, that relatively few words were examined (Saragi et al., 1978; Dupuy & Krashen, 1993; Horst et al., 1998; Rott, 1999a). It seems difficult to give a robust statement on how effectively reading promotes incidental word acquisition, without incorporating a large number of target words into reading studies. Schmitt (2010) argues that:

[T]he time required to learn a word in a small set of lexis may be much less than the time required to learn the same word in a larger set of vocabulary, due to the possible decreasing efficiency of whatever learning strategy is being used. It would thus be erroneous to generalize the faster rate of acquisition from a small set of words to a much larger set of words. (p. 165)

Another problem with the reading tasks used is found in the study by Dupuy and Krashen (1993), despite it being widely cited in the literature because of the substantial acquisition reported by their subjects. By viewing a video of five scenes from the story, subjects in this study were further exposed to the target words, raising the question of whether incidental learning in this case should be attributed solely to the reading task. In Rott's (1999) study, the texts were created with sufficient contextual clues to enable the learners to infer the meaning of the target words. This procedure, however, is not typical of incidental vocabulary learning from reading. The short paragraphs could have been very easy for the subjects to understand, thus facilitating their inference of meaning from the context. This might not have been the case if other texts, such as graded books, were used. An additional problem, concerned with the reading of the texts, with the previous research is found in the study by Horst et al (1998). The fact that the teacher read the text and the subjects followed him/her along in their books, raises the question of the effect of the additional aural input on the amount of incidental gain. However, in the study, this gain was attributed mainly to reading. In another study, it appeared reading while listening led to better incidental learning than in the unaccompanied reading condition (Brown et al., 2008). Moreover, although such an approach enabled Horst et al. to ensure that all subjects were exposed to the entire text, it is possible that the participants' reading of the book was disrupted by the

teacher's reading out loud. Evidence from similar research shows that sustained silent reading helps subjects to recognize unknown words and remember them in vocabulary tests (Day et al., 1991).

A further serious criticism of these studies, as Nation (2001) points out, concerns the lack of control over the level of difficulty of the reading materials. For instance, Saragi et al. (1978) and Pitts et al. (1989) assigned their subjects to read 'A Clockwork Orange', but failed to take into account the length and difficulty of the novel. Pitts et al (1998), reported that, due to the difficult language found within the two chapters, half of their subjects could not complete them in the assigned time.

In an attempt to eliminate these research flaws, a study by Waring and Takaki (2003) applied a different testing approach to assess incidental vocabulary gain from the 'A Little Princess' graded book and to examine the long-term retention of newly-acquired words. In this study, the researchers asked 15 intermediate volunteer female Japanese students to read and enjoy the graded book, which was at a lower level than their subject language. The purpose of providing an easy text, as Waring and Takaki explained, was to ensure that the words surrounding the target words were familiar enough to assist in guessing the latter's meaning. To select the words to be tested, Waring and Takaki (2003) devised a new technique: they chose 25 words with varying repetitions (1 to 18 occurrences) in the text and then replaced them with disguised forms. The substituted items looked like English words, but had no meaning. For instance, the word "house" was replaced with "windle" and "window" with "bettle". The substituted items served to ensure that all target words were unfamiliar to the subjects, therefore, any reported gain of these words could be ascribed solely to reading.

Unlike many previous L2 studies, which used a multiple-choice test as the only testing instrument, Waring and Takaki (2003) were the first to develop a

methodology for small gain measurement through multiple testing formats: a word form recognition test, a meaning (translation) test, and a multiple-choice recognition test. Their study sought to determine whether words that appear more frequently in a reading text are more likely to be learned and retained, rather than forgotten over time.

Moreover, this study investigated the relation between the number of repetitions of a word and long-term retention. Whereas most studies of the incidental acquisition of vocabulary are based on a word meaning measurement, given immediately after reading (Pitts et al., 1989; Day et al., 1991; Dupuy & Krashen, 1993; Horst et al., 1998), to my knowledge, Waring and Takaki (2003) is one of only a few studies to investigate word retention over time. The subjects were tested three times over a period of three months. First, the researchers administered the three test formats to the subjects immediately after they had finished reading. The same three tests were given to the subjects one week later and then again in a delayed test after three months.

Significant results appeared in the two immediate post-tests (word form recognition and multiple-choice recognition). From the 25 target words, the subjects obtained mean scores of 15.3 for correct spelling ($SD=4.4$), and the mean score for meaning recognition by multiple-choice test was 10.6 ($SD=4.0$). However, the meaning translation test yielded a substantially lower mean score 4.6 ($SD=3.5$). The repetition of words in the text notably affected the word gain results in the immediate post-test. The results obtained from post and delayed tests in this study indicate that the incidental learning of words declines with time. The study also shows that there is a strong correlation between word frequency in a text and its retention. A word-form appearing more than eight times had a 50% chance of being remembered by the subjects after three months. However, this was not the case when the repetition effect on word-form meaning recognition was examined. Waring and Takaki (2003) reported that L2 learners may need to meet the word more than 20 times to

incidentally learn its meaning from reading. The researchers concluded that, even if a word was encountered more than 18 times, the chance of retaining its meaning was only between 10 and 15 per cent after three months. In all three tests, there was no instance of words being retained if they occurred fewer than eight times.

Waring and Takaki (2003) found an overall decline in the mean scores of all tests over the three months, and, that relatively few learned words were remembered over the period. In fact, they discovered that, on average, the meaning of only one out of the 25 words was correctly recognized by their subjects three months after reading. Based on their results, the researchers concluded that the amount of incidental vocabulary gained from reading depends on the type of tests used to measure the gains. They also found that reading one graded book could be relatively useful in helping the learners to improve and enrich knowledge of words they already know.

Compared to the previous studies, Waring and Takaki's research seems to follow more rigorous methodological procedures. One striking advantage in this study was that, in the instrumental testing design and presentation, Waring and Takaki (2003) used more sensitive tests, by administrating three vocabulary testing formats that all tested the same words three times over three months. The word meaning translation test served as a supplement, providing more information as to whether answers in the multiple-choice test were the result of the subjects making inferences, or of actual learning of the test words. An interesting finding, with regard to the other two testing formats (i.e. word form recognition and multiple-choice recognition), indicated that the translation test used in this study was the most demanding test, as it involved unassisted recognition of a word meaning. This finding may lead us to question the rates of learning reported in studies using multiple-choice tests as a main or sole research instrument (Saragi et al., 1978; Pitts et al., 1989; Day et al., 1991; Dupuy & Krashen, 1993; Rott, 1999a). Waring and Nation (2004, p. 16)

argue that “being able to choose an appropriate meaning from a list of plausible choices, as in the multiple-choice test, shows that at least some knowledge of form and meaning has been retained, even though, in many cases, it may not be enough for unassisted recall”. More interestingly, Waring and Takaki were the first researchers to measure retention of incidental learning over time. The three-month delay test provided a clearer picture of the actual learning that remained in the learners’ minds. The decline of word knowledge over time found in this study also leads to further questions about previous findings from immediate post-tests (Dupuy & Krashen, 1993).

Having said that, however, Waring and Takaki’s study (2003) also suffered from a number of limitations. Although the researchers considered learning in their study as gain from a kind of ER, they were still unable to address the reading conditions recommended for ER, particularly, the learners’ ability to self-select books that suit their language level and interests. Like previous studies pointed out earlier (Saragi et al., 1978; Day et al., 1991; Horst et al., 1998), subjects in Waring and Takaki’s study read only one graded book, which was chosen by the researchers. More seriously, the graded book used by Waring and Takaki (2003) was very easy and below the subjects’ proficiency level. If Krashen’s 'Comprehensible Input Hypothesis' $i+1$, which suggests that input leads to better acquisition when it is slightly above the current competence level of the learner, is considered here, then reading, as a means of incidentally improving vocabulary knowledge, was unlikely to be effective in this study. A similar weak point is related to the fact that the subjects who participated in this study were volunteers. Since they volunteered to take part in Waring and Takaki’s (2003) study, these learners may have been highly motivated and therefore, not representative of the global population. Another issue that needs some consideration is the use of made up words as the objects of learning. Is it ethically appropriate to require learners to learn such words? Were they aware of the fact that they were not real words and did this affect their incidental learning in some way?

A recent case study of an adult Japanese ESL learner in the USA was carried out by Furuta (2012). Similar to Waring and Takai (2003), incidental learning in this study was seen by the researcher as taking place through extensive reading, irrespective of the fact that only 21 pages were given to the subject. Target words in this study were identified by the learner. The researcher asked the subject to highlight any unknown vocabulary she came across while reading, including the words she looked up in a dictionary, and annotate them in the margins. In total, there were 116 words in the text that were found to be unfamiliar to the subject. Among these words, 36 were looked up in the dictionary. In order to assess the incidental gains from reading, Furuta (2012) used a self-report checklist containing a list of all the 116 items. The test was given to the subject one week after she had finished the reading. She had to rate her knowledge of a word from three choices: 1) YES if the subject knew the meaning of the word; NS (not sure) if the subject had some idea of the meaning but was not quite sure if it was the right answer; and NO if the subject did not know the meaning of the word. The NS option was included to register any partial knowledge of the target words.

Findings from this case study suggest that reading can significantly enhance L2 incidental vocabulary acquisition. According to the data, about a third (35%) of the target words were acquired by the Japanese subject, after reading two chapters of a novel. In total, 41 words out of the 116 were rated as YES by the subject. Because the subject was an ESL learner in the USA, and it was supposed that she would have had exposure to the target words outside the text, the researcher took a further step to check whether the learning had occurred solely from the reading. Furuta (2012) used online software to analyse and categorise the target words according to their frequency of use in English. In order to make a more reliable statement about the possible gains from reading, the researcher looked at words from a low frequency band, on the basis that such words had little chance of being met through the input the subject was exposed to in her everyday life. Furuta found that 24 of the 41 words his subject acquired belonged to the off-list, therefore leading to a safe conclusion

that about 20% of 116 total words had been learnt. This means that even when highly common words were excluded from this study, the results still showed significant incidental gain in comparison with earlier studies, which reported on a far smaller number of words.

However, like the other studies reviewed above, this study was not problem-free, as there were several methodological limitations that could flaw its findings. First, the learning reported on by the subject was based on her self-reported indication of knowledge. Selecting 'Yes' (I know the word meaning) did not necessarily mean that the subject really recognised the correct meaning of that word. The picture of actual learning would have been clearer if Furuta (2012) had asked his subject to provide translations or synonyms for the words she rated 'Yes'. Moreover, since the subject was asked to highlight unknown words, it is very possible that she paid some degree of attention to these words, which could have affected the incidental acquisition status in this study. Similarly, the subject was permitted to use a dictionary and therefore, the learning reported cannot be solely attributed to the incidental acquisition mode.

To sum up, early research using short readings to examine incidental acquisition of words suffered from a number of methodological flaws. The low rate of gains reported in most of the studies discussed so far could be attributed to research design factors, such as lack of sensitive testing instruments and the small number of reading texts. Recent studies, measuring incidental vocabulary gains from ER over considerably longer periods than in studies described so far, have been more interesting, as they have made brave attempts to address the flaws associated with the previous incidental vocabulary learning research.

3.3.2 Incidental vocabulary learning from ER

Up to this point, the most popular L2 studies dealing with incidental word acquisition from short readings have been reviewed. Nowadays, researchers have become well aware of the importance of ER in developing language

learners' level of competence. Thus, a recent trend has been towards the use of an ER approach where learners read different texts, typically through graded readers (see Day & Bamford, 2002).

Cho and Krashen (1994) introduced what they called 'free reading' to four female, international students living in the United States. The participants were asked to read books from the 'Sweet Valley' series (popular novels written for young American readers) for pleasure. In order to identify the tested words, the researchers asked the subjects to underline any new words they encountered while reading the books, but no further information about the aim of the study or the vocabulary test was given. The number of titles read by each of the four subjects varied between eight and 23, while the time required to finish reading ranged from two weeks to two months. Based on the words that each subject underlined, Cho and Krashen (1994) asked three of them to write synonyms or a translation in their first language, while the fourth participant was asked to give an oral definition in English of the words consistently underlined by the others.

This study revealed significant gains in the subjects' vocabulary knowledge. The highest score attained by one of the subjects was 80% of the target words encountered in the story. The lowest word gain score obtained by another subject was 56%. However, despite the fact that Cho and Krashen's (1994) experiment is one of only a few ER studies to examine learners' acquisition of a large number of new words, there remains a question concerning their research procedures and methodology. One of the study's design problems lies in the choice of reading material. One of the subjects was able to read 23 books in less than one month, or approximately six books a week. In view of her having been able to read so much in that period of time, it is possible that the text was too easy. Equally, as Laufer and Hulstijn (2001) caution, it cannot be guaranteed that the vocabulary was learned incidentally from repeated exposures while reading, since the dictionary usage in this study could have

allowed for some intentional learning. The study showed that the two subjects who consistently looked up the meaning of unknown words in a dictionary while reading learned 17 and 34 items in each book, whereas the subject who did not consult the dictionary learned only seven words. Likewise, given that the experiment took place in the USA, it is likely that subjects had additional encounters with the target words outside the reading materials. As a result, it is unlikely that learning could be entirely attributed to reading. In fact, it is difficult to tell, as the researchers did not attach the test or the words to their article.

More recent research of incidental vocabulary learning from ER has used a better-designed methodology. Pigada and Schmitt (2006) were the first to look at the learning of different aspects of word knowledge: meaning (form-meaning relationship), form (spelling), and grammatical behaviour (use). In this study, a Greek native speaker was offered one month of ER, before incidental acquisition of the target words was measured. The subject was a French language learner whose proficiency level was “lower intermediate”. For the reading, the ‘Lectures CLE en Français facile’ graded readers (Level 1) were used. From a collection of 17 books, the subject chose titles that were of the most interest to him, and read roughly one reader each week. In this study, a one-on-one interview with the participant took place at two different stages (before and after reading), in order to examine his knowledge of 133 target words. They consisted of 70 nouns and 63 verbs, and were divided into six frequency groups (1, 2-3, 4-5, 6-10, 10+, 20+ occurrences). The subject was not aware of the aim of the study or of the words being tested, nor was he told that he would have a post-test that was exactly the same as the pre-test.

One of the interesting techniques used by Pigada and Schmitt (2006) was the distraction task. Because both pre-and post-test were taken over two days according to the test type (a spelling test on the first day and meaning and grammatical behaviour on the second day), the subject was given an irrelevant

task to perform after finishing the spelling test. Such a task would have “served to flush the target words from his [the participant’s] immediate memory” (2006, p. 11), so that the researcher could guarantee that no additional direct learning of these words would take place after the session. Moreover, in order to avoid parallelisms in the test form, the word order was changed in the meaning and grammatical behaviour tests, which were taken simultaneously the next day.

According to the results of this study, the subject showed a 65% improvement in at least one of the knowledge aspects of the target words. Spelling knowledge improved significantly, from about 37% before reading to 60% in the post-test. However, it was found that knowledge of nouns was better enhanced for all six frequency groups in comparison with knowledge of verbs, which increased for all but the 6-10 and 20+ verb occurrences groups. The meaning knowledge was also enhanced. The subject recognized the meaning of 8% of the words in the interview held before he began the extensive reading, while the post-test revealed word meaning attainment for about 24% of the target words. Overall, word meaning knowledge was significantly enhanced in all frequency groups; but there was no learning of verbs that appeared only once in the texts, and only very slight learning was reported for nouns occurring only once or 2-3 times. Similarly, the participant’s grammatical mastery showed improvement when he was tested after reading. His grammatical knowledge of nouns was enhanced in all frequency groups, from about 13% to 43%. Improvement was also observed in his grammatical mastery of verbs, which reached 20%, as compared with 4% in the pre-test. This improvement in verb mastery occurred in all frequency groups except the 10+ group. Examining the 133 target words individually, Pigada and Schmitt reported that understanding of about half of them (49.6%) improved in one knowledge aspect, 10% in two knowledge types, and 7% in all three knowledge aspects.

Although the study by Pigada and Schmitt is one of the few which created a plausible environment for learning vocabulary from extensive reading, some problematic issues are associated with their research. Given the fact that the participant was a motivated French learner, incidental gains of vocabulary might have been exaggerated in this study. One might wonder whether less highly motivated subjects could reach the same figures. Moreover, since, as the researchers indicated, their Greek subject spoke English fluently, it is possible that his previous experience of learning a foreign language could have improved his ability to learn words incidentally from reading in French. Research in second language learning shows that language skills are likely to be transferable from one foreign language to another (Larsen-Freeman and Long, 1991). Thus, readers with a very limited L2 learning experience may not find it so easy to acquire so much vocabulary from reading extensively over a period of one month.

In a further study, Brown et al. (2008) examined incidental vocabulary learning within three extensive modes of learning: reading, reading while listening, and listening. This experiment involved 35 adult Japanese EFL learners, who were randomly divided into three groups (12 subjects in group A, 14 in group B, and 9 in group C). The subjects' English language proficiency level was either pre-intermediate or intermediate. In this study, Brown et al. (2008) used three high-beginner level books, from the 'Oxford Bookworms' graded readers, which had not been read or seen by their subjects. The researchers made their selection from this level because they assumed that, if their subjects knew 96% to 99% of the running words, it would be easy for them to successfully guess the meaning of unknown words from the context. Twenty-eight words were selected from each book (the total number of target words being 84) to be tested in the three experimental modes. Selection of the target words was based on their occurrences in the books. Seven words were chosen from each of four frequency bands (2-3, 7-9, 10-13, and 15-20) tested in this study. However, due to their high frequency usage level, to avoid the risk of being already known by the subjects, these 84 target items were replaced by unreal words (similarly to

the technique in Waring and Takaki (2003). For example, the word “happy” was changed to “mird”, “mouth” to “reak”, and “bag” to “slape”.

To determine which of the three extensive inputs would better enhance the incidental acquisition of the words in the four frequency bands, subjects in each group experienced the three stories in different modes. In other words, if the subjects in group A experienced story 1 in reading mode, story 2 in listening mode, and story 3 in reading-while-listening mode, the subjects in each of the other two groups would encounter these stories in different modes from those of group A. This meant that, by the end of the treatments, each story would have been used once in every mode of input. For the treatments in which the reading alone and reading-while-listening input conditions were examined, pleasurable reading, as well as reading and listening, was promoted. Before the two treatments began, the subjects were provided with a short written introduction to each book, which served to provide schematic background for the three stories. Moreover, each story contained eight words from outside the target frequency range. Thus, definitions of these words were made available to the subjects to maintain the consistency of coverage rate. For the listening-only treatment, the subjects listen to recorded versions of the stories, which were read at a very low speed by one of the study researchers. In addition, short written introductions to these stories, plus a set of pictures from the original books, were provided in this treatment. The pictures were offered to assist the subject in following the stories' narration. To relieve the burden of a long listening task, the subjects were given a three-to-four-minute interval in the middle of the session. In these three experimental modes, the subjects were not allowed to consult a dictionary, but were told to guess, as far as possible, the meanings of the unknown words.

Like Waring and Takaki (2003), Brown et al. (2008) used multiple testing instruments to investigate incidental learning and retention of words over time. They administered two types of tests: the multiple-choice test (recognition

test), which included the correct answer, three distracting words, and the “I don’t know” option; and the meaning-translation test, which was used to measure meaning recall. In order to measure learning and retention rates, the researchers administered the two tests three times: immediately after the experiments, one week later, and three months later. To avoid any transfer of knowledge from one test type when answering the other, the two test types were administered separately, with the translation test being given before the multiple-choice test. One point was given for correct answers in both test formats; but for measuring partial word knowledge in the translation test, a half point was given if the participants provided similar Japanese meanings. To collect data from each input mode, the subjects in each group were tested immediately after finishing whichever input mode they had experienced in the first week of the experiment. For instance, a test was administered to group A after they had finished reading; to group B after they had finished listening; and to group C following their reading-while- listening task. After an interval of one week the same tests were taken by each group. Then, the same two-phase testing procedure (immediate and one-week-interval tests) was implemented after the participants in each group had experienced a different input mode, in the fourth week of the study. However, in the sixth week, following each group’s final experimental input, the subjects were only given immediate vocabulary tests. After a week, the three groups were asked to write a short essay about their impressions of the experiment, including their feelings about the stories and their preferred input mode.

Brown et al. (2008) found that incidental vocabulary learning was demonstrated by the immediate post-tests for all three input modes. The amount of learning, however, varied from one input mode to another. Mean scores averaged for the two test types showed that both reading-while-listening and reading alone obtained higher results in comparison with listening alone. On recognition tests, out of 28 words, the 35 subjects had a mean average of 13.31 for reading-while-listening, 12.54 for reading, and 8.20 for the listening alone sessions. Similarly, the recall tests on the three experimental modes gave

mean scores of 4.39 for reading-while-listening, 4.10 for reading alone, and 0.56 for the listening mode. When comparing the scores for the three input modes over the three test periods, Brown et al. noticed that, for all three input modes, there was a small decline in the mean scores obtained in the recognition test (multiple-choice test). In addition, for the three test administrations, there was an overall mean score decline for the three modes of input, when scores from the recall test (meaning translation test) were compared. The researchers found that only one word was recalled from the reading-while-listening and reading alone modes in the three-month delay test, but no words were learned from the listening alone input. Findings from the two test types in this study are consistent with the aforementioned conclusion of Waring and Takaki (2003), in which they argued that incidental vocabulary learning measurement is largely affected by the type of testing instrument.

Although Brown et al. (2008) stated that their research focused on incidental word learning from three extensive inputs, in fact, the way the study was implemented did not represent extensive learning conditions. The researchers used three books, but, by contrast with extensive reading, the participants in the three groups of this study read only two titles (one for reading and another for reading-while-listening), and only listened to the third. Even in the case of the two titles read by each group, there remains a question as to whether learning from the book each group took in the reading-while-listening experiment should be considered as learning from a solely ER input. Moreover, in determining the target words, Brown et al. (2008) selected a set of 28 words from each of the three books. Thus, the amount of incidental vocabulary learning reported in this study, particularly the one word gain in the three-month delayed test, was a result of reading one book and not three books as the researchers stated. Once again, the use of made up words might have affected retention rates.

More recently, Al-Homoud and Schmitt (2009) provided a more realistic extensive input in their study. The researchers studied the effects of ER on a number of language skills, including incidental learning of vocabulary. They compared the vocabulary growth of students receiving two different reading approaches: extensive and intensive. In this study, the researchers invited 70 EFL learners who were studying in one of the Saudi universities to participate. Generally, as the researchers themselves affirmed, the participants' language proficiency was relatively weak and they were novice learners. However, their levels were diverse, as some of the participants' language competence apparently exceeded that of the others.

The participants were randomly divided into an intensive group (23 students) and an extensive group (47 students). The intensive group took a reading course, given by an instructor who was a colleague of one of the researchers. The participants took four 50-minute reading classes during each week of the experiment. In the course of the ten-week session, they read short texts from a book called 'Reading Power'. Every week, the intensive group was given explicit teaching of 80 to 120 new words from texts read with the teacher in the class. Other activities related to vocabulary learning and reading strategies were also provided in the intensive reading classroom. Overall, the group read about 100 pages over the experimental period. After every session, the intensive-approach group was required to perform additional reading tasks and to answer comprehension questions at home.

In contrast, the participants in the extensive group received a 10-week reading programme, during which they were exposed to 150 graded books with different titles. These books represented various genres and were selected from different publishers: Oxford University Press, Longman, and Cambridge University Press. In order to match the participants' interests and degrees of language proficiency, the books also varied in subject matter and language levels (from Level 1, 300-400 Headwords, to Level 6, 3000 Headwords). A

small mobile library of six shelves in a wheeled bookcase was created. To help the participants borrow books suited to their language proficiency, the 150 readers were labelled in six different colours according to their level of difficulty and each level was assigned to one of the shelves. The extensive group took their reading sessions in the classroom. Under the supervision of one of the researchers, the library was brought to the classroom during every reading session, so that participants could choose titles of interest. Because this study investigated other language learning skills alongside vocabulary learning, the extensive group was further divided randomly into a vocabulary group (21 students) and a fluency group (26 students). Based on their existing reading ability, as determined by the pre-test, the vocabulary group was asked to read books which were one level beyond their existing reading level, while the fluency group read titles one level below their ability. Through this design, the researchers wanted to ensure that the experiment contained enough new unknown words for the vocabulary group to encounter, as well as sufficiently easy texts for the fluency group. During the programme, the participants read various amounts of text, with those who read books from the higher levels reading more pages every week than those who selected lower level books.

Al-Homoud and Schmitt (2009) noticed that, regardless of the graded reader levels the participants chose, they could all read one book per week after three weeks of ER classes. The researchers estimated that, by the end of the treatment, the readers of the higher level books could read approximately 162.000 words, whereas participants reading low level readers could read considerably fewer, or about 31.500 words in all reading sessions.

In order to determine the amount of vocabulary knowledge growth of the two groups of participants, two versions of the Vocabulary Levels Test (see Schmitt et al., 2001) were given them both before and after the treatment. This test, originally designed to estimate the vocabulary size of English language learners, consists of five sections based on word frequency levels, namely

2000, 3000, 5000, the Academic Word List, and 10,000 words. However, due to the participants' low proficiency level, Al-Homoud and Schmitt only tested three frequency levels (in the 2000, 3000, and Academic Word List sections). Every level consisted of ten clusters and every cluster contained six words with three definitions. The participants were required to match the definitions with three of the six words in the cluster. To measure the vocabulary knowledge improvement at the three tested levels, Al-Homoud and Schmitt compared the results for the 23 intensive group participants with those of the 21 extensive vocabulary group participants, before and after the ER course had finished.

The researchers found that, as with its effects on other learning variables, ER facilitated the participants' incidental learning of a similar amount of new vocabulary to that acquired by the intensive reading group, who received explicit instruction based on the words. Both extensive and intensive groups improved their vocabulary knowledge on all three frequency levels to a statistically significant degree ($p < .001$), but there was no significant difference in the extent of word improvement between the two groups ($p > .05$). The researchers also concluded that the greater the frequency level of words, the more word learning chances participants in both groups had. They estimated that, from the 2000 level, the extensive group learned 5.57 words, while the intensive group learned 6.57 words per day. For each day of the study, the two groups, reported gains of 2.00 and 2.03 words from the 3000 level and 0.46 and 0.62 words from the Academic Word List, respectively.

Despite the fact that Al-Homoud and Schmitt provided a wide range of readers, as is typically required for an ER programme (Day & Bamford, 1998), their findings should be viewed with care. The report of significant word learning by the two groups in response to different reading approaches was based on estimates of changes in the participants' vocabulary size. Al-Homoud and Schmitt did not indicate how many of the newly learned words really occurred in the graded books that their participants read. Horst (2005, p. 362) points out

that “word knowledge measures administered before and after an ER treatment should feature words that participants have actually met in their reading”. Moreover, Al-Homoud and Schmitt failed to eliminate the suspicious role of other L2 input sources found in previous research (see for example, Cho & Krashen, 1994). Their study examined the learning of words at 2000, 3000, and Academic Word List levels. Given that the participants were university students taking an intensive EFL course, it is likely that they had learned many of the words at these three levels through other classes in their course. The findings would be more reliable if Al-Homoud and Schmitt had pre- and post-tested the participants on a set of words selected from the readers, along with a control group used to observe the effect of outside sources.

The best-designed study to date was conducted by Horst (2005). In fact, the current research design stems from Horst’s study, and is heavily based on her methodological limitations. In her pilot study, Horst investigated the incidental vocabulary enhancement of 21 EFL learners with different mother-tongue backgrounds. Before beginning the six-week treatment, the participants had been studying English in Canada for various periods of time. Their proficiency levels ranged from elementary to upper intermediate. Alongside their EFL course, the participants were involved in 6 weeks of extensive reading. In this course, a large number of graded readers, at six levels of simplification, were offered. Horst created a mini-library of 70 titles written in different genres, and provided two copies of each title.

In most previous research, with participants reading the same text, certain sets of words were tested before and after reading, to assess the amount of new word knowledge acquired. In Horst’s experiment, such a process was not feasible since each participant would have had to read a set of books that differed from those selected by the other participants. Horst, thus, was the first to use individualized tests for her participants’ pre and post vocabulary learning measurement. First, she scanned 12 readers electronically, using two

books from each level of simplification. Because the scanning of entire books was found to take a great deal of time (more than three hours for books of over 100 pages), Horst decided to scan only the first 20 pages of each. Using lexical frequency profiling software, Horst created 12 lists of word types occurring in the 12 books (1000, 2000, off-list, and Academic Word List). For the pre-reading vocabulary knowledge measurement, Horst used a modified version of the test that she and a colleague had used in an earlier study (Horst & Meara, 1999). The self-report checklist pre-test included 100 words sampled randomly from 12 readers: 50 items from the 1001-2000 most frequent words list and 50 off-list words. Horst excluded words from the 1000-words list and the Academic Word List from the test because she assumed that 1000-level words would be familiar to her participants, while scanning produced few words from the Academic Word List.

The checklist pre-test was administered to the participants before they read the material and required the subjects to choose one of three options: YES, NOT SURE, and NO. The description for these three options is provided in the previous section (see Furuta's 2012 study). By including the (NOT SURE) option in the pre-test, the researcher wanted to register any partial knowledge of the target words' meaning.

After taking the test, the participants started the six-week ER program. During the treatment period, the library was made conveniently accessible during breaks from the participants' regular ESL classes, and they had the opportunity to check out books that were interesting and suitable for their levels of proficiency. In addition to their free reading at break time, they were allowed to borrow their self-selected books. Horst reported that the participants did most of the reading independently at home. To promote extensive reading, the researcher provided one hour of reading activities every week, including paired discussion of the readers, and reading and vocabulary tasks. Horst did a 20-page scan for each new title selected by the participants during the experiment

period. Only 62 titles, which were found to have been borrowed by more than two participants, were targeted. Horst was eventually able to scan 37 of these titles, which meant that 159 of 222 books were checked out during the six-week experiment. For the individualized post-test design, the researcher created lists of words appearing in the 37 titles. Every individualized list contained 100 words taken equally from the 1001-2000 and off-list frequency ranges. Each individualized post-test was created entirely from items occurring in four of the titles read by the participants during the six weeks. Seventeen of the participants satisfied this criterion and so took the individualized post-test with the three options: YES, NOT SURE, and NO.

In order to assess any growth in vocabulary knowledge, Horst compared the number of words that a participant had rated YES on the pre- and post-tests. The results indicated a significant post-test mean knowledge increase in both the 1001-2000 most common list ($M = 6.59$, $SD = 5.47$) and items from the off-list frequency range ($M = 10.29$, $SD = 7.62$). This represented a new gain of seven words from the 1001-2000 frequency band, and ten words from the off-list items. In total, Horst found that about half of the 35 words rated NO in both frequency bands in the pre-test were learned by her participants following the ER treatment.

Nevertheless, Horst argued that her results for off-list word gains might be more reliable, as other sources of exposure could have contributed to the learning of words from the 1000-2000 most frequent words zone. In order to further examine her participants' new off-list vocabulary knowledge, as it appeared in the checklist post-test, the researcher asked them to take another individualized post-test. This time the Vocabulary Knowledge Scale test (by, Wesche & Paribakht, 1996) was administered. The test only included words which a participant had rated NO (unknown meaning) in the baseline pre-test and, which occurred in the books s/he had read during the treatment. According to these criteria, 16 of the 17 participants who took the first post-test were able

to take the Vocabulary Knowledge Scale test. However, Horst found that only 35 words could qualify for this test, and the number of words each individualized test included varied between one to three items. Findings from this test again showed new partial or full growth in knowledge of 18 words, representing 51% of the items targeted for this test.

Horst (2005), however, could not avoid a number of methodological weaknesses. Because of being unable to entirely scan the 70 books, she selected the target words from the first 20 pages of each story. This range of course does not represent all words in the books. Another problem is that Horst gave her subjects two different checklists as pre- and post-tests to measure their incidental vocabulary gains from the extensive reading. Such a testing technique could have affected the degree of gain reported in her study. Rather than examining word knowledge improvement through two different target word lists, she could have asked participants to select the books they were interested in reading during the six weeks in advance. This step would have allowed the researcher to identify the target words from the subjects' preferred titles in order to create the same individualized pre- and post-tests. In addition, there was a risk that the use of a checklist test design could have made the results less reliable, by enabling the participants to engage in guesswork. Paired discussion and reading activities provided during the experiment could also have brought the subjects' attention to the target words.

Still, in comparison with other studies of incidental L2 vocabulary learning discussed in this chapter, Horst's findings seem very encouraging, particularly if we consider her better-implemented methodology. There are several factors which distinguish Horst's study from related research. First, her experiment took place in a relatively typical and effective ER environment (Day & Bamford, 2002). Furthermore, the amount of reading Horst offered in her research and the number of words tested exceeded those used by any other researcher. Moreover, Horst was the first to attempt to individualize the

instruments for testing her participants. She also used computer programs in order to determine and classify the actual words encountered by her participants in a very large number of texts. Another striking factor is that her test design allowed subjects to register the partial knowledge of words, taking into account the incremental nature of incidental vocabulary learning (Nagy & Herman, 1985).

Having discussed the research on incidental vocabulary and problems associated with the reading studies, it is obvious that the research has been marred by a number of methodological limitations. Although most recent studies of reading have applied more well-designed methodologies in measuring incidental learning of L2 vocabulary, these studies failed to avoid several problems that were associated with earlier research on incidental L2 vocabulary acquisition. From the above review of these studies, it seems that Horst (2005) has offered the best designed study to date, as her study seems to have been carried out with minimal methodological problems. The present study, therefore, replicates most of the methods used by Horst, while aiming to avoid the limitations found in her study and in previous research. This strategy will be discussed in the methodology chapter. Major limitations that will be addressed in this study are summarised as following:

- 1) none of the studies measured pre-post gains of words that learners encountered in various numbers of ER texts;
- 2) in general, a very small number of words were measured in incidental vocabulary studies;
- 3) word learning reported in these studies was mainly measured in terms of receptive knowledge through testing word meaning acquisition and very little consideration was given to the knowledge of words at the productive level;
- 4) with the exception of one study (Waring & Takaki, 2003), which had a 3-month delayed post-test for measuring word retention over time, most

studies provide a ‘snapshot’ of incidental learning, through immediate post reading tests and;

- 5) previous studies did not draw a clear picture of the amount of individual vocabulary learning that occurred when participants had read different texts extensively that they perceived to be at their level of proficiency, and of interest to them. Although a few attempts (e.g., Cho & Krashen, 1994; Horst, 2005) have been made in this regard, the researchers, as discussed above, failed to apply an accurate measurement of word knowledge development from the reading materials.

Chapter 4:

Methodological considerations and pilot study

Chapter 4: Methodological considerations and pilot study

4.1 Introduction

This chapter discusses the methodological issues involved in planning and executing the experimental research. It begins with a discussion of the underpinning ontological, epistemological, and methodological considerations that motivated the selection of the research design and implementation of the current study. The chapter also presents the objectives of this study, the research hypotheses and research questions before it explains the ethical considerations. The final section describes the procedures of implementing the pilot study and summarises its major findings.

4.2 Ontological and epistemological perspectives

Ontology refers to perspectives of reality. It is a discipline of philosophy, which deals with the study of what exists, or, in other words, whether there is truth to be discovered. (Crotty, 1998). Ontological assumptions are concerned with how individuals perceive social reality. The fundamental points of orientation here are the two distinguishable ontological positions: objectivism (positivism) and constructionism (interpretivism). Objectivism “is an ontological position that asserts that social phenomena and their meanings have an existence that is independent of social actors” (Bryman, 2004, p. 16). From the positivist’s ontological point of view (realist), there is a single external reality. That is, researchers see reality as external, independent and objective, regardless of their beliefs. The positivist view is that reality exists ‘out there,’ and the subject under analysis should be tested objectively (Remenyi et al., 1998). In contrast, Constructivism (interpretivism) is a position “that asserts that social phenomena and their meanings are continually being accomplished by social actors” (Bryman, 2004, p. 17). Constructivists are concerned with the

belief that “the observer makes a difference to the observed and that reality is a human construct” (Wellington, 2000, p. 16). They believe that there is no single reality, the world has different meanings and knowledge is personal, experiential, and subjective. Researchers adopting this approach ask questions of their subjects and make their judgements based on human perception, which is subjective and socially constructed according to its context (Opie, 2004; Cohen et al., 2008). In other words, different researchers can obtain different conclusions from the same enquiry depending on their ontological perspective (Crotty, 1998).

Epistemology is the study of knowledge (Crotty, 1998). It is “the philosophy of knowledge, a system of thoughts that articulates specific beliefs about the nature of knowledge” (Hartas, 2010, p. 16). In other words, as Blaikie (2000) points out, epistemology is concerned with the central issues and questions of ‘truth’ including, “what can be counted as knowledge, what can be known, and what criteria such knowledge must satisfy in order to be called knowledge rather than beliefs” (p. 8). In social science research, similarly to ontology, there are two common epistemological positions: objectivism and constructivism (Crotty, 1998). The distinction between quantitative and qualitative approaches to research is due to the different underlying epistemological positions held by researchers within these two paradigms. Objectivists hold the view that knowledge is objective, and independent of the observer. They view knowledge as real ‘hard’ and objective and require observation, measurement, quantification, and generalisation of data (Wellington, 2000). Constructivists, on the other hand, believe that reality is subjective and socially constructed. Their view of human knowledge contradicts the objectivist view. The constructivist theory is based on the assumption that reality and the investigator who observes it cannot be separated. They believe that the investigator does not discover truth, but that s/he constructs it, and, as a consequence reality can have more than one interpretation (Crotty, 1998; Wellington, 2000).

4.3 Research methodology

Methodology refers to the “theory of getting knowledge”. It is concerned with the specific strategies and methods a researcher uses to obtain knowledge about social reality (Hartas, 2010). It connotes a set of principles and rules that guide an inquiry, and the particular methods and procedures that are developed in order to understand or deal with a problem (Crotty, 1998, p. 16). More simply, methodology is concerned with the use of the most appropriate methods or procedures (techniques used for data collection and analysis), in order to provide an evidential basis for the construction of knowledge about a phenomenon being investigated (Opie, 2004). It is stressed that methodology should not only provide a rationale for the choice of particular methods, but it should also justify the reason for the employment of particular strategies within these methods (Crotty, 1998).

In the social sciences, there are two approaches to research: deductive and inductive. The former is associated with the positivist paradigm, while the latter is linked the constructivist approach. According to the positivist methodology, the researcher deductively establishes hypotheses to test theory from already existing knowledge, by using quantitative methods (Duffy, 1985). Quantitative research requires the researcher to collect numeric data through research tools such as questionnaires or multiple choice questions. Then, the researcher uses specific statistical techniques to analyse these data, in order to prove or disprove a hypothesis. Researchers who adopt quantitative methods use them to make generalisations from a sample to a larger population (Ivankova & Creswell, 2009). Inversely, a constructivist methodology requires the researcher to inductively generate a new theory, which evolves from the data, through the implementation of qualitative methods (Duffy, 1985). In qualitative research the researcher gathers his/her data through interviews, observations, or images. The aim is not to prove or disprove a phenomenon, but rather to explore and portray the phenomenon that is being examined (Ivankova & Creswell, 2009).

4.4 Selecting and designing paradigms

Every piece of social research or enquiry is guided by a set of beliefs, which is known as a paradigm. At a very early stage in an enquiry, the researcher needs to make assumptions about human knowledge and about the nature of reality s/he will encounter in her/his research. Accordingly, these assumptions determine the way the researcher understands her/his research questions, the research methods s/he uses, and how the researcher interprets her/his research outcomes (Crotty, 1998). Crotty (1998) argues that researchers need to adopt a philosophical paradigm to inform their methodology. Saunders et al. (2009) specify that the philosophical branches that every researcher needs to ponder are related to ontological, epistemological, and methodological expectations. Huberman and Miles (1998, p. 181) stress that it is a ‘healthy medicine’ for researchers to be aware of the philosophical expectations that guide their research.

The research design adopted in this study is largely influenced by Grotjahn’s (1987) classification of research paradigms in the L2 acquisition arena. According to him, the distinction between the qualitative and quantitative traditions in Applied Linguistics has been oversimplified, and research should be considered from three different angles: (1) the design or the method of data collection (whether it is experimental or non-experimental), (2) the type of data (whether it is qualitative or quantitative), and (3) the sort of data analysis (whether the analysis procedures entail statistical or interpretative analysis). In line with this distinction, Grotjahn (1987) suggests eight research paradigms. In order to answer the research questions stated below in section 4.7, this study used a mixed design, as is proposed by Grotjahn’s (1987) eighth paradigm (Experimental-Quantitative-Interpretive) as shown in Figure 4-1 below. According to this paradigm, research is conducted using both experimental methods, which entail quantitative data and statistical analysis, and qualitative data collection and its interpretation. Although the quantitative approach will

be the leading approach for data collection and analysis, a qualitative method will be utilised to enrich the data for the fourth research question.

Figure 4-1: Types of research design in the language learning field (from Grotjahn, 1987)

PURE FORMS

Paradigm 1: exploratory-interpretive

- 1 non-experimental design
- 2 qualitative data
- 3 interpretive analysis

Paradigm 2: analytical-nomological

- 1 experimental or quasi-experimental design
- 2 quantitative data
- 3 statistical analysis

MIXED FORMS

Paradigm 3: experimental-qualitative-interpretative

- 1 experimental or quasi-experimental design
- 2 qualitative data
- 3 interpretive analysis

Paradigm 4: experimental-qualitative-statistical

- 1 experimental or quasi-experimental design
- 2 qualitative data
- 3 statistical analysis

Paradigm 5: exploratory-qualitative-statistical

- 1 non-experimental design
- 2 qualitative data
- 3 statistical analysis

Paradigm 6: exploratory-quantitative-statistical

- 1 non-experimental design
- 2 quantitative data
- 3 statistical analysis

Paradigm 7: exploratory-quantitative-interpretive

- 1 non-experimental design
- 2 quantitative data
- 3 interpretive analysis

Paradigm 8: experimental-quantitative-interpretive

- 1 experimental or quasi-experimental design
 - 2 quantitative data
 - 3 interpretive analysis
-

Cohen et al. (2008) note that the kinds of research questions that researchers address are controlled by their different ontological and epistemological positions. My research poses questions with causal explanations, which seek both an objective/positivist and subjective/interpretative understanding of the phenomenon. Given the vital aim of the present study, which is to investigate incidental EFL vocabulary learning from ER, I have employed the process-product paradigm (Shuell, 1996), which is concerned with how the process affects the product. This, as mentioned above, was achieved through adaptation of Experimental-Quantitative-Interpretive paradigm suggested by Grotjahn (1987). In L2 learning research, the experimental strategy has been one of the most powerful methods for discovering the strength of relationships between different variables. In classroom-based studies, researchers have typically explored the effect of innovative teaching methods or materials on students' language learning (Nunan, 1992). This method is clearly implied in this definition of an experiment, which describes it as:

[a] test or series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reasons for changes that may be observed in the output response (Montgomery, 2005, p. 1).

The current study aims at exploring the cause-effect relationships between two variables: ER and L2 vocabulary development. Therefore, it is believed that the choice of the experimental design will allow me to maintain a control over the complex processes involved in the implementation of this enquiry, and the measurement of its outcomes. From the epistemological point of view of the objectivist, the reality investigated in this study is independent from the researcher and therefore it is believed that knowledge about the world can be acquired by objectively gathering facts in a systematic way. This can essentially be achieved, by implementing an experimental strategy and testing the hypotheses, with the purpose of building laws that would be applicable at a global level.

Although the objectivist paradigm is the leading approach, as most of research questions require the testing of hypotheses, the interpretivist approach is also involved in this study, as it also employs a non-experimental design, involving interviews with the participants. The aim of this is to gather qualitative data about the learners' vocabulary learning experience of and their perception of the ER programme. In this respect, the forms of data collection and analysis used are largely dependent on how knowledge is constructed.

4.5 Objectives of the study

The objectives of this research study are to:

- determine the amount of incidental vocabulary acquired by Libyan adult EFL learners, who extensively read different self-selected graded materials that suit their language proficiency levels;
- determine the number of times words need to be met in ER materials in order for incidental learning to occur at the levels of both receptive and productive knowledge;
- investigate the extent to which words gained incidentally from ER can be recalled by Libyan learners over specific periods of time;
- investigate the extent to which individualisation in selecting reading materials leads to any significant difference in the number of words learned incidentally by learners, and to gain understanding of the perceptions of the adult, Libyan, EFL learners' concerning the ER programme.

4.6 Research hypotheses

- There will be a significant difference in the vocabulary knowledge scores of learners exposed to the ER treatment as compared with the scores of those who did not participate in the ER programme.

- There will be a significant difference in the number of times a word needs to be met in graded readers, in order to be acquired incidentally at both receptive and productive levels.
- There will be a significant difference between the amount of acquired vocabulary in the post-test and a 9-month delayed test.
- Individualisation in selecting and reading different texts from a collection of a variety of graded books will not reveal a significant difference in terms of the amount of vocabulary the participants obtain incidentally from these materials or their perception of the ER programme.

4.7 Research questions

Question 1: Does reading extensively enhance adult, Libyan EFL learners' vocabulary knowledge?

To what extent can a two-month ER programme enhance incidental learning of English vocabulary items by adult, Libyan, EFL learners for receptive and productive use?

Question 2: Does word repetition affect receptive and productive incidental vocabulary learning from extensive reading? If so, how many exposures are needed for incidental vocabulary learning?

Question 3: How much of the newly learnt vocabulary is retained 9 months after the ER programme?

Question 4: Will individualisation in the selection of ER texts reveal any significant difference in the number of vocabulary items learned, as well as in the learners' perception of the ER programme?

4.8 Ethical considerations

In social research, ethics is an important consideration. The main principle of research ethics is that the participants should be protected from any mental or

physical harm which might occur as a result of being involved in an investigation (Wellington, 2000; Dörnyei, 2007; Bryman, 2008). In view of this, a number of educational organisations, such as the British Psychological Society and the British Research Association, have offered ethical codes to researchers in the field of social sciences. My research complies with British Educational Research Association (BERA) guidelines, including: informed consent, anonymity and confidentiality. Punch (1994, p. 90) makes it clear that “a professional code of ethics is beneficial as a guideline that alerts researchers to the ethical dimensions of their work, particularly prior to entry”. In this regard, several ethical issues were considered during the implementation of this research and they will be considered throughout the research stages. It has been stressed that “ethical concerns should be at the forefront of any research project and should continue through to the write-up and dissemination stages” (Wellington, 2000).

According to the BERA guidelines, any social research requires the obtainment of the institution and the participants’ agreement to assist in an investigation. Prior to conducting my research, the permission of the institution was obtained from the English Language Department. It permitted me to use the departmental documentation and rooms, and to invite the students to take part in this experimental research.

Since my research involved adult, Libyan EFL learners, it was necessary for me to obtain their consent. Before that, I provided the participants with a full oral and written explanation of my research in Arabic language. In this, I informed them of all areas of my research and clarified any aspects that might reasonably be expected to influence their willingness to participate in the experiment. I also answered honestly any questions raised by the students regarding ER in general, my research procedure, and the participants’ role in particular. Once the participants showed their agreement to take part in the research, they were asked to sign the consent form. This also meant a

confirmation that the participants had been provided with sufficient information about the research procedures and their obligation towards them. It also indicated that all possible questions about the research had all been fully answered by the researcher.

Nonetheless, due to the sensitivity of incidental learning of vocabulary as a research topic, I limited the participants' understanding of what the research was about. In doing so, I believed that the participants would respond more naturally to the experimental treatment. The study could have been invalidated if the participants had been informed that they were involved in research of incidental vocabulary acquisition. As Fraenkel and Wallen (2006) state, in particular instances, social research cannot be carried out unless some deception of the subjects takes place. Therefore, the information given to the participants was not complete. It included the general benefits that a language learner can gain from participating in an ER programme, without mention of the incidental gain of vocabulary. According to Anderson and Arsenault (1998), researchers can implement a 'deception' experiment if it is the only way to discover the truth, and providing that no harm comes to the participants.

An additional ethical issue was related to the research sampling. As a random selection of participants was used in this research, I was aware of the fact that their participation was voluntary and that some students might be unwilling to be involved. The consent process, therefore, respected and protected the right of the participants' self-determination without coercion. Participants had the right to refuse to participate in the research or to withdraw from participating at any point in the experimental period without repercussions.

Another ethical concern relevant to this research was the confidentiality of the data. Participants were required to put their names on all data collection forms. However, to achieve confidentiality and anonymity, it was guaranteed in the research information form that the participants' information, including personal

or any other identifying details, would not to be used in any publications. All participants were assured that any data collected from or about them would be kept confidential. Therefore, participants' names were removed from all data collection forms and they were coded by letters and numbers.

In any social research the potential benefits must outweigh the burden of subjects' participation and cooperation in the enquiry (Anderson & Arsenault, 1998). In my opinion, considering the issue of harm and benefit in my research, the Libyan EFL learners could be regarded as benefitting from the research, as reading long interesting texts in a relaxing environment could provide them with additional English language learning input.

In order to maintain a high level of enthusiasm during the experiment, refreshments were provided to participants free of charge. Likewise, a friendly, relaxed atmosphere was created by providing the research library with additional furniture, such as sofas, tables, and chairs. Although there was no promise of personal benefits, understanding the need to 'give something back' to the research subjects was my primary concern. In this regard, small gifts were offered to all participants at the end of the experiment. Letters of appreciation were also given to them, expressing my gratitude for their participation during the experimental period. Moreover, two copies of each title of the 40 research books were given to my research participants. They were invited to select the titles they were interested in, to keep for themselves.

At the research preparation stage, many students, including control group participants, attended discussion sessions on the ER program, but, due to the research design, they were not involved in the reading treatment. For ethical reasons, I donated a set of all 40 titles of the research books to the English language department. These books were made available to interested students who had missed the reading program, as well as other colleagues in the English language department.

4.9 Pilot study

Before the main study was conducted, a pilot study was carried out in the same department, to identify the appropriate levels for materials and to test the procedures of the ER programme, as well as the adequacy of the data collection instruments. This pilot study also aimed at testing the participants' pre-existing knowledge of the target words from the high frequency band and ensured that the items from this frequency were not familiar to them. Moreover, it attempted to find deficiencies in the design of the proposed experiment or procedure, so they could be addressed before the beginning of the main study. Mackey and Gass (2005, p. 43) suggest that a pilot study is "an important means of assessing the feasibility and usefulness of the data collection methods and making any necessary revisions before they are used with the research participants". Thomas (1998) insists that typical piloting procedures should involve five key aspects:

- a) selection of subjects who share similar characteristics with those who will participate in the research project;
- b) pilot-study subjects should be informed that their feedback will help in refining the main study procedures, materials, and instruments;
- c) administration of the research instruments to the pilot subjects;
- d) elicitation of information about problems which the pilot subjects may have encountered with piloting materials, instruments, and procedures;
- e) revision of the research procedures, materials, instruments on the basis of the pilot-study results.

The level of graded books to be used in the experimental study was identified through several piloting procedures (see section 4.9.2 below). Likewise, the target words belonging to the second thousand frequently used words in English (2K) were also determined through a pilot checklist test. As for checking the feasibility of using a relatively large number of lexical items in the pre- and post-tests, it was important to pilot the amount of time the

participants in the main study would need, to respond to the various numbers of individualised test words.

4.9.1 Pilot subjects

The subjects selected for the pilot study were EFL students from the same department as the participants for the main study. Because of the sensitivity of the research topic, the pilot subjects were told that they were participating in an assessment programme for language learning and teaching in their department. Though the pilot subjects were not targeted in the main study, they were not informed about the aim of the pilot study, as this may have led them to link it to the later experimental research. The majority of the pilot subjects had different roles: either in determining book levels (40 students) or in the process of identifying the target words from the 2k frequency band (34 students). Six subjects, from those who took part in piloting the target levels of graded books, were further involved in reading, testing and checking the clarity of the self-reporting instruments. To ensure an equal representation of subjects in the main study, two students were randomly selected, according to their choice of books from each of the three target reading levels.

4.9.2 Piloting stages and results

4.9.2.1 Determining levels of reading materials

Firstly, the levels of graded books that the participants in the main ER study would be reading were determined. Three months before the main study began, three titles from each of the seven graded levels (by Oxford University Press, Table 4-1) were purchased for the purposes of piloting. These books were provided by the same publisher of the books that would be used in the main study. In total, a collection of 21 books was introduced to 40 students. These students were randomly selected as representative of the whole target

population in the four academic year classes (10 students from each academic year).

Table 4-1: Bookworms Graded Scheme (Oxford University Press)

Stages	Headwords	Average word count	CEF Level
Starter	250	950-1635	A1
Stage 1	400	5200	A1/A2
Stage 2	700	6500	A2/B1
Stage 3	1000	10000	B1
Stage 4	1400	16000	B1/B2
Stage 5	1800	23000	B2
Stage 6	2500	30000	B2/C1

Headwords: they are words that form headings in a dictionary, under which their meaning is explained

CEF⁴ Level: A1=Beginner, A2=Pre-intermediate, B1=Intermediate, B2=Upper-intermediate, C1= Advanced level.

To determine which of the seven stages was likely to be most suitable for the students in the main study, the 40 subjects were asked to read some pages from the seven graded levels and decide the levels of reading materials which were suitable for their proficiency level. As for the construct validity of the subjects' self-selected levels of reading, each subject was further tested on the level of book s/he had selected. The book levels tests used for this purpose are offered by the publisher of graded books and are available online⁵ (see Appendix 1). According to the publisher, the level is considered appropriate for the reader if s/he obtains scores between 80% and 95%. Most of the subjects obtained the pass mark for the level of reading they selected. If they did not, however, students were given another chance to select more appropriate levels, which were also confirmed by the book level test.

⁴ CEF: Common European Framework Reference for Language

⁵ Available at: <https://elt.oup.com/student/bookwormsleveltest/?cc=us&selLanguage=en>

Levels of the graded books/ main findings

The pilot study showed that certain levels of the Oxford Bookworms Books would be suitable for the participants in the main study. The vast majority of the 40 pilot subjects (39 subjects = 97.5%) found books at stages 3, 4, and 5 suitable for their language proficiency level. The pilot subjects' self-reported book levels were further checked through the Bookworms Reading Level Tests, for which each participant was given a test that matched her/his selected level. There were several cases (7/40=17.5%) in which disagreement between the subject's self-reported reading level and the required score on the corresponding level test was found. Therefore, the subjects were asked to reselect more appropriate reading levels confirmed by the same testing procedures. This finding suggested that, checking participants' choice of reading levels would be an essential step in the main study and that providing certain support for participants, as to how to identify their correct individual reading levels, would also be important. Therefore, the same validation process was followed in the main study to ensure that the participants in the ER programme would be reading with sufficient comprehension.

4.9.2.2 Determining target 2K words

Two word frequency categories occurring in the reading materials were used in the main study: the 2000 most frequent word families in English (this word frequency will be alternately referred to as 2K in this study), and low frequency word families (alternately referred to as off-list words). The latter category of words was assumed to have little chance of being encountered outside the graded books, therefore, there was no need to pilot them in this study. However, because of their high frequency of use in the English language, it was necessary to ensure that the 2K target words used for assessing incidental vocabulary acquisition in the main study were not already known by the participants.

To this end, an online software called Vocabprofile⁶ (Cobb, 2010) was run. The VocabProfile tool design assists analysis of word frequency in the English language. It calculates the proportions of low and high frequency words used by a native speaker in a particular written text. By counting the frequency of both words and word families, VocabProfile classifies the words in the text into four frequency bands (1K = the 1000 most frequent word families, 2K= the second most frequent 1000 of English, AWL= academic word list, Off-list words= words, which are not included in the aforementioned bands, and proper nouns).

A Corpus analysis of the entire set of books selected for use in the main study produced a considerable number of words belonging to the 2K zone, at each of the determined levels. For instance, the output illustrated that the 2k words represented 5.82% (516 word families) of the entire number of texts at stage 3, about 6.27% (633 word families) at stage 4, and 6.20% (686 word families) of at stage 5. A sample of the Vocabprofile output of 2K word families in stage 5 books is shown in Table 4-2.

Table 4-2: Sample of 2K word families (Stage5 books)

2k families: [686:1563:14610]

pale_[39] pan_[2] parcel_[4] pardon_[3] parent_[30] park_[66]
passage_[10] passenger_[26] path_[41] patient_[20] pattern_[3]
pause_[32] pearl_[2] peculiar_[1] pen_[9] pencil_[2]
perfect_[60] persuade_[41] pet_[1] photograph_[11] pick_[55]
pile_[10] pin_[1] pink_[12] pipe_[15] pity_[27] plane_[20]
plaster_[5] plate_[13] plenty_[15] pocket_[26] poem_[9]
poison_[7] police_[89] polish_[1] polite_[56] pool_[4] pot_[25]
pour_[17] practical_[5] practise_[17] praise_[9] pray_[10]
precious_[1] prefer_[20] pretend_[40] priest_[13] print_[3]
prison_[38] prize_[1] probable_[56] profession_[10]

⁶ Available at: <http://www.lexutor.ca/vp/eng/>

Because of the large number of 2K words occurring in the books at each target stage, it was not possible to create a test that included all of the words in the books at a particular stage, as this might overburden the test takers. Therefore, it was more feasible to create several tests for the words at each stage. Words in these tests were randomly selected, through the use of Microsoft Excel software. Because the testing format (checklist test) allowed measurement of a large number of target words, it was decided to include 100 items in each test. This produced 5 lists for stage 3 and 6 lists each for stages 4 and 5 (see Table 4-3).

Table 4-3: The number of checklist tests at each stage

Book level	No of 2K word families	No of checklist tests	Average of items in each checklist test
Stage3	516	5	103 or +1
Stage4	633	6	105 or +1
Stage5	686	6	114 or +1

A checklist test format was used to obtain a reliable estimate of the global knowledge of these words. This test was originally created by (Horst & Meara, 1999) and was modified later by Horst (2005). Two pilot subjects took one checklist test at each stage. In total, 34 subjects answered 17 different checklist tests. As can be seen in Table 4-4 below, the test included three options: ‘YES’ (I know the meaning of the word), ‘NS’ (not sure of the word meaning), and ‘NO’ (I don’t know the word meaning). As stressed by Horst (2005), including the ‘NS’ option would be of twofold benefit: firstly the subject could report their partial knowledge of a word and secondly, the option would allow them to be more honest with their self-reported answers.

Table 4-4: Self-reported checklist test Sample 2K items

44- parcel	YES	NS	NO
45- hide	YES	NS	NO
56- stretc	YES	NS	NO
47- drawer	YES	NS	NO
48- curl	YES	NS	NO
51- row	YES	NS	NO
52- rent	YES	NS	NO
53- shame	YES	NS	NO
54- breath	YES	NS	NO
55- hunt	YES	NS	NO
56- combine	YES	NS	NO
57- intend	YES	NS	NO
58- stove	YES	NS	NO
59- priest	YES	NS	NO
60- stupid	YES	NS	NO
61- suspect	YES	NS	NO
62- rival	YES	NS	NO
63- temper	YES	NS	NO
64- patient	YES	NS	NO
65- practise	YES	NS	NO

The selection of 34 test takers was made on the basis of their scores on the Oxford Bookworms Level Tests. With the cooperation of their teachers, the reading Level Tests (stage 3, 4, and 5) were given randomly to 100 students, across different classes in the department (25 from each academic year). This aimed at gaining enough possible representatives at each of the three reading stages. From this number, it was possible to randomly select 10 students from stage 3, and an equal number of 12 students from stages 4, and 5. As stated above, each two pairs of subjects took one checklist test. If a word obtained an interrater agreement of ‘NO’, then this word was assumed to be unlikely to be known by the participants and therefore qualified for testing in the main study.

Target 2K words/ main findings

Overall, 1835 lexical items from the 2K categories that appeared in the books used in the main study, were tested in this pilot. The major purpose of this pilot test was to ensure that items from this high frequency band would not be familiar to the main participants. As stated above, words rated ‘NO’ by both subjects who took the same checklist test qualified for inclusion in the main study. Less than half of the 2K words in the entire set of texts of the main study met this criteria. Based on the subjects’ interrater agreement, 683 words out of

the total number of tested 2K items (1835 words) were rated as unknown by pilot subjects. Table 4-5 categorises these words according to the stage of the books.

Table 4-5: Results of pilot 2K words

Book level	No of 2K word families	Words rated 'NO' both checklist takers	%
Stage3	516	211	40.89
Stage4	633	279	44.08
Stage5	686	193	28.13
Total	<i>1835</i>	<i>683</i>	<i>37.22</i>

4.9.2.3 Piloting reading programme and vocabulary testing procedures

This piloting programme was carried out one month before the main study. It aimed at checking aspects related to the main experimental work, including the pre- and post-tests and the subjects' reading experience. Six subjects (2 male, 4 female) were randomly selected from the 40 students, who, at an earlier stage of the pilot study, participated in determining the target book levels (see section 4.9.2.1 above). In considering the variation of the levels of target readers, two subjects were randomly selected to represent one of the three reading stages. The materials used for the main study (40 titles of graded books), were presented to them and they were encouraged to select one book at their already-identified reading level. The pilot testing instruments and reading programme procedures were conducted in three stages as follows:

Before they started reading, individual lists of target test words were created from the corpus data of the six books selected by each subject. Based on the entire number of words in a book at each of the three stages (see Table 4-1), one word out of every 500 words was chosen to be tested. This resulted in an equal number of 20 words tested by each of the two pilot subjects who read the stage 3 books, 32 words for every pair who read a book at stage 4, and 48

words for each of the two who read the stage 5 books. In all these individualized tests, one third of the items included were from the 2K category, and two thirds were from the off-list category. Three days before they started their reading, the six subjects were pretested on these words. Using the testing instrument which would be used in the main study, the subjects had their individualised target words listed on a four-scale test. The written instructions on the test were also provided orally, to ensure that subjects fully understood how to answer the test (see Table 4-6).

Table 4-6: Individualised test sheet

Below you have 4 choices. You need to choose one answer for each word in the table. Write A, B, C, or D to show your knowledge levels of a word next to it in "Answer" column. <u>If you chose answer (D) use the word in a sentence.</u>		
A. I don't know the word.		
B. I have seen the word before, but I am not sure of the meaning.		
C. I understand the word when I see it in a sentence, but I don't know how to use it in my own writing.		
D. I know this word and can use it in my own writing.		
Word	Answer	Use the word in a sentence if you choose answer Level 'D'

Once the pilot reading program had started, the library was made accessible to the pilot subjects during the university opening hours. Under the researcher's supervision, a similar environment to that in the main study was created in order to pilot reading and testing procedures. After finishing their selected books, the participants were requested to fill in an instant book report (see Appendix 2). The report sought to find out the amount of actual reading carried out by the participants, as well as asking them to rate their comprehension of, and their comprehension difficulties with the books they selected themselves. Three days after they had finished reading, the same individualised pre-vocabulary tests were given to the six subjects, for the purpose of piloting the post-reading vocabulary learning assessment. #

Reading and vocabulary testing / main findings

The pre-and post-tests in this pilot study revealed that there was no serious concern about the language difficulty of the test instructions, as it was further explained orally. Therefore, the same procedures were applied in the main study to ensure the validity of the participants' answers. The pilot study, nevertheless, proved to be useful in revealing that the test instrument needed some modifications. The major concern about the scale was related to some equivocal answers provided by subjects at levels C and D. For example, words such as *herb*, *canal*, *refugee*, and *midwife* were rated by some pilot subjects at level 'C' (receptive knowledge) in the pilot scale, but it was not possible to determine whether these words were really known receptively by the subjects or whether the level was rated arbitrarily. Similarly, it was difficult to judge some ambiguous sentences such as '*I saw anchor*' at full knowledge level 'D'. Therefore, it was felt to be safer to add another column to the testing instrument when it was used in the main study. This column would ask the participants to provide an equivalent Arabic translation of the words they rated at 'C' or 'D' levels. Providing an Arabic translation for the word would serve to reduce the possibility of the word being arbitrarily reported as receptively or productively known.

The pilot subjects were required to complete time records, which allowed the researcher to identify the time spent on reading by each of them. They were also asked to report their reading experience through the instant book reports that they filled in after finishing their books. The pilot study proved that the instant book report was adequate in terms of the presentation and the clarity of its questions and therefore no modification. In terms of the readability of the target books, it was found that only four of the six books used in the pilot study gained over 90% rate of comprehension. It was important, therefore, to take an additional step to ensure participants in the main study would choose books at the appropriate levels, by using the level tests provided by the publisher.

In terms of the amount of time taken for reading, the records showed that the time spent on reading one graded book varied between the six subjects and that reading one book took between seventeen to twenty-two hours approximately (6 to 9 days), regardless of the level of the book (see Table 4-7). This was an encouraging rate for successful reading in the main study, particularly, as the pilot reading did not largely differ from previous research findings in terms of quantity, which suggests that for an effective ER programme, students should be reading one book per week or more at their proficiency level (Nation and Wang, 1999).

Table 4-7: Piloting the time required for reading a graded book

Piloting Subjects	Booh level	Title of the book	Approximate hours required for finishing the book	Reading days	Average of reading hours per day
Subject 1	Stage 3	Skyjack	18 hours	6 days	3.0
Subject 2	Stage 3	Marthin Luther King	22 hours	9 days	2.4
Subject 3	Stage 4	Three Men in a Boat	19 hours	8 days	2.4
Subject 4	Stage 4	A Tale of Two Cities	20 hours	8 days	2.5
Subject 5	Stage 5	The Bride Price	19 hours	9 days	2.1
Subject 6	Stage 5	The Dead of Jericho	17 hours	8 days	2.1

In terms of the vocabulary tests, the pilot study revealed further concerns. The 6 individualised tests included items that appeared in one book. Nevertheless, an overall estimation indicated that the subjects needed a maximum of 11 minutes to answer 20 items. This led to a consideration of the amount of time that would be needed if the scale tests included larger numbers of lexical items selected from several books. Therefore, the pilot testing criteria, based on selecting one word in each 500 words the subjects read, were considered problematic for the main study participants, since they would be reading much longer texts over a longer period of time. For instance, one pilot book at stage 5 produced 48 words. If the participants in the main study chose to read 5 books at the same level then this would result in a total of 240. This, according to time estimated in this pilot, would require the subjects to spend more than two

hours answering questions on such a large number of words. This, unfortunately, may have led the subjects to experience a moment where they felt fatigued and bored with the long test. Therefore, I decided to reduce the target words by targeting one in each 1500 words that appeared in the texts that main subjects read. This was believed to be a more practical approach, in terms of the number of test items used in the main study.

The one third inclusion of 2K word families was found to be an appropriate portion, in order to compensate for the scarcity of words from the low-frequency word family band. More importantly, a very small number of 2K words were rated as known at “C” or “D” levels by the participants (12/ 64= 18.75%), leaving a very hopeful impression about the effectiveness of the piloting procedures involved in determining which words from this high frequency band could be used in the main study.

To sum up, this pilot study revealed useful information which helped to revise the research procedures, materials, and instruments for the main study. The next chapter covers the adapted, final intervention programme. It provides a detailed description of the design, preparation, and application of the ER programme, including a discussion of the phases of data gathering.

Chapter 5:

Research design (main study)

Chapter 5: Research design (main study)

5.1 Introduction

Chapter 5 describes the methods involved in investigating the effect of ER on incidental L2 vocabulary learning in a Libyan EFL context. It provides a detailed account of the design and implementation of this experimental research. First, the rationale for choosing the research method is briefly stated, then, a discussion of the research context is offered. A description of the participants and the procedures involved in selecting the research group is also offered in this chapter. Finally, the stages in which this experimental work was carried out and the instruments used to collect the data are discussed.

5.2 The experimental design in detail

Nunan (1992) states that experiments are generally conducted to discover the strength of relationships between two or more variables, such as the relationship between innovative teaching materials (an independent variable) and students' test scores (a dependent variable). The current study was developed as an extension of an experimental pilot study (by Horst, 2005) that explored the effects of ER intervention on ESL learners' incidental vocabulary acquisition. With reference to reading research reviewed earlier in the literature review chapter, this experimental strategy has been very commonly used in measuring vocabulary acquisition from reading and listening input.

In his review of L2 research methodology and design, Nunan (1992) identifies three types of experiments: the pre-experimental design, the quasi-experimental design, and truly experimental design (see Table 5-1 below).

Table 5-1: Types of experiments (Nunan, 1992; p. 41)

<i>Type</i>	<i>Characteristics</i>
Pre-experiment	May have pre-and posttreatment tests, but lacks a control group
Quasi-experiment	Has both pre-and posttests and experimental and control groups, but no random assignment of subjects
True-experiment	Has both pre-and posttests, and experimental and control groups, and random assignment of subjects

These three types of experiments vary in terms of randomisation of samples and the use of control groups. This study strongly favoured the truly experimental design research over the other two types of design for two main reasons. First, if no random assignment of subjects had been applied, as a quasi-experimental design suggests, then a larger population would have been required. Taking into account the very controlled design of this investigation, involving a large number of participants would have required unfeasible amounts of research groundwork, including a larger quantity of text scanning and corpus work. Likewise, because of the overall cost and time, it was not feasible to carry out such an experimental design with non-random subject assignment. One common practice to address the challenge of using the whole target population in a particular piece of research is through the random selection method. Saunders et al. (2009) argue that a random sampling technique is essential when:

- it would be impracticable for you to survey the entire population;
- your budget constraints prevent you from surveying the entire population;
- your time constraints prevent you from surveying the entire population;
- you have collected all the data, but need the results quickly. (p: 212)

Therefore, in order to avoid these challenges, the truly-experimental design was selected for carrying out the current study. It was an experimental case study which investigated in detail, the incidental learning of vocabulary of a particular group within the whole EFL population of eight Libyan state universities.

The use of the case study method has been applied in various research areas and seeks “to contribute to our knowledge of individual, group, organizational, social, political, and related phenomena” (Yin, 2009, p. 4). Nunan and Bailey (2009, p. 158) define a case study as “a detailed, often longitudinal, investigation of a single individual or entity (or a few individuals or entities)”. Yin (1994, p. 13) describes a case study as an empirical inquiry that “investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident”. As pointed out by Mackey and Gass (2005), case studies, in general, aim to offer a holistic explanation of language learning or use within a particular situation and population. The present study looks at the incidental learning of vocabulary demonstrated by Libyan, EFL students in one of eight state universities. A detailed description of this experimental case study is provided below.

5.2.1 Research context

As stated earlier in section 1.4.2, Libyan tertiary education is provided by eight public universities located across the country’s regions (Gulinao, 2013) and education in these universities is free. Before being enrolled in the tertiary stage of education, students have to finish three educational stages: six years in primary education (age 6 to 12), 3 years in preparatory education (age 12 to 15), and 3 years in secondary education (age 15 to 18). At the age of 18 students in Libya start their university studies, for which they are enrolled in a variety of specialisations. This research was conducted at one of the eight Libyan public universities belonging to the Ministry of Higher Education. The

role of the university where this research was conducted, is to accept undergraduate students who have finished their secondary school education in any of the Libyan regions. The university has nine different schools: Arts, Science, Medicine, Economics and Accountancy, Engineering, Law, Agriculture, and Physical Education. The English Language department (the focus of this research) is one of the School of Arts' departments. Most of the students who graduate from this department are assigned as English language teachers in public secondary schools.

5.2.2 Participant description

Mackey and Gass (2005) claim that the number of participants needed for L2 research ultimately depends on the type of the study conducted. As they note:

These research types can range from an intensive experiment including several treatments, pretests, immediate posttests, and multiple delayed posttests, all entailing complex and finely grained linguistic analyses, to a large-scale second language testing study, in which simple numerical before and after scores may be utilized for hundreds of learners (p. 124).

In a similar vein, Fraenkel and Wallen (2006) propose a sample size guideline, which suggests minimum numbers for three types of research in education. They claim that an approximate number of 15 to 30 participants can be appropriate in experimental research which requires maintaining control of the effect of several variables. In contrast, a larger number of about 100 participants or more may be necessary for descriptive research and 50 participants for carrying out correlational research. Based on this categorisation, this study fits into the first category, as it adopts an experimental research strategy. The implementation of this study involved very complex design and procedures, which made it unfeasible to involve a very large number of subjects.

The number of participants involved in this research was originally 80 Libyans (some withdrew at an advanced stage of the research), who were native

speakers of Arabic (23 males and 57 females). They were divided equally into experimental and control groups. The participants were EFL learners in the English Language Department. They represented students from all of the department's classes and shared very similar English language learning experiences and backgrounds. The ages of the students ranged from 18 to 22 years, and all of them received the same amount of instruction every week. On average, they attended university six days a week and they took three 2-hour EFL classes each day. In addition, they studied in the same EFL learning environment, in which exposure to and use of English outside the classroom were limited. These participants started learning English as a mandatory subject from the 7th grade of the preparatory school (i.e., between the ages of 12 and 15). After they had finished the 3-year preparatory stage, they had studied three years in secondary schools specialising in English (i.e., between the ages of 15 and 18). In these schools, English was taught intensively as a major subject and students were required to attend three 45-minute classes a day, including lessons on writing, reading, speaking, listening skills, and lab work (Asker, 2011).

In general, it was assumed that the language proficiency level of students in the English department was between intermediate and upper-intermediate level. The pilot subjects who randomly represented the students from the research site were able to read books at levels A2 and B2, which, according the CEF Levels' description, represent levels between pre-intermediate and upper-intermediate (see Table 4-1 in section 4.9.2.1)

5.2.3 Participant selection

In view of the purpose of my research, the overall approach adopted was to randomly assign participants who represented all of the language levels at the research site to the experimental and control groups. Two main sampling strategies, probability and non-probability sampling, are discussed by a number of researchers (Wellington, 1996; Saunders et al., 2012). The former strategy

involves a random selection of the sample, while the latter is based on the subjective judgement of the researcher. In view of the purpose of my research, the overall approach adopted was to randomly assign participants who represented all language levels in the research site. One major advantage of using probability sampling is that, compared to the non-probability strategy, it reduces any risk of bias and subjects with different levels of ability have an equal probability of being selected (Cohen et al., 2008).

Before selecting the research participants, there was concern about the inclusion of the pilot subjects in the main study. This was because they could have experienced some of the main research materials and procedures. Given the sensitive nature of incidental vocabulary learning assessment, it was thought necessary to exclude them from the main study. Table 5-2 illustrates the numbers of students enrolled in all four-years of the department's courses during 2010/2011 academic year.

Table 5-2: The number of four academic year students in the English department

First year	Second year	Third year	Fourth year
83	77	61	68

Several procedures were applied in the process of selection of the 40 experimental participants and their 40 corresponding partners in the control group. First, a research subject pool was put together, involving 150 students. These students were chosen from various English Language department classes, across the four year course. It should be mentioned here that 19 students from each academic year participated in piloting the levels of reading materials, the target 2K words, and the reading treatment and testing procedures (discussed above in section 4.9.2), so these students were excluded from the present study. By using the department database, the ID numbers of all students in the department (i.e. 64 students in the first year; 58 in the second year; 42 the third year; 49 in the fourth year) were utilised to produce a random

number of the 150 subjects. They were selected by running a simple Random Sample function in Microsoft Excel. In order to ensure, as much as possible, an equal distribution of these subjects from across the department's four years, a random choice of 38 students was made equally from each of the first and second years (as their number was larger than in the other two advanced years) and 37 students were equally chosen from each of the third and fourth year classes, providing 150 potential participants in all. It was assumed that having a similar number of students from each year would offer an equal opportunity for students from all four years to be chosen, when the random selection of the experimental and control groups was applied.

Because this study employed an experimental and control group in exploring the effect of the ER treatment on incidental vocabulary growth, it was important to ensure that participants in these two groups had a similar vocabulary knowledge levels before they were pre-tested. To this end, the Vocabulary Levels Test (VLT), which has been used in a large number of vocabulary studies (e.g., Cobb, 1997; Schmitt & Meara, 1997; Laufer & Paribakht, 1998; Al-Homoud, 2007), was applied. The VLT version 1 (Appendix 3), developed by Schmitt (2000), was used to give an estimate of all of the participants' current vocabulary sizes. This test was properly validated by using a range of techniques for analysis (see, Schmitt et al., 2001). It includes five sections, which are divided according to word frequency levels (2000, 3000, 5000, 10,000 and the Academic word List, AWL), each of which has ten clusters (i.e. 30 items). Since words belonging to the AWL band were not targeted in the present study, they were removed from the test. Therefore, students were only required to answer the remaining four test sections.

With the cooperation of two of my colleagues, instructions on how to answer the test were given to the students. Without mentioning the purpose of the test, these students were told that it was given to them as part of general evaluation of their department courses and it was made clear to them that the test results

would have no effect on their grades. All of the 150 students were then invited to sit the VLT. If they agreed, they were then asked to sign their consent. In order to remain anonymous, students were asked to write their student IDs, and unwilling participants were free to withdraw. All students took the test at different times depending on their university class timetables. The researcher and the two colleagues interchangeably administered the test. Students who either failed to finish the four frequency levels sections, or did not report their student IDs were excluded from the study. This left 139 students who satisfied these two conditions and who were targeted as the population for the present research.

So as to reduce any possibility of the VLT being linked to the research programme, a two-week interval was left between the test and the introduction to ER approach. During this period, the ER library was prepared. Before selecting the research sample from the target population, all the students were provided with information about the ER research programme. Given that this reading approach is not currently used in teaching and learning English in Libya, it was important then to ensure that participants were fully aware of its nature. My major concern was to convince the students that ER was to their benefit. Therefore, the entire number of participants were invited to attend a meeting in one of the university lecture halls to introduce them to the ER programme. The session lasted for about two hours, during which time, the students who attended had the opportunity to participate in a discussion. A detailed explanation of every aspect of the programme was offered. The emphasis during the discussion was particularly placed on the role that reading interesting and comprehensible stories plays in improving the general language competence of second language learners. Because incidental vocabulary knowledge measurement is a highly sensitive topic, the students were not informed about the main aim of the research. Instead, they were told about the general benefits that could be obtained from participating in the reading programme.

Despite the majority of students attending the meeting, it was considered necessary to provide additional sessions to ensure that all the target students had been exposed to this new reading approach. Consequently, I set up several short, optional discussion sessions in one of the departmental rooms. Students attended in small groups and if a student joined one of these sessions, s/he was not required to participate in another. Students who missed the chance of participating in the first main meeting were provided with the same information about the ER approach. Some of the students attended the main meeting, but for more enquiries and clarification about the programme, they also joined in the short discussions. During these discussions, I explained the concept of ER and how it differed from their traditional intensive reading in class. This included the key points of ER (e.g. reading for pleasure in one's free time and reading at one's language level). The research procedures and the participants' role in the reading program were also clarified. Again, to ensure the research aim remained hidden, there was no reference to the incidental learning of vocabulary. At the end of the discussion, 12 students said that they were not enthusiastic about taking part in the research, so they were omitted from the study.

For the research design purpose, participants in the experimental group were selected randomly by using a simple random sample function in Microsoft Excel, which was run using the students' ID numbers to produce a random number of ten students in each academic year. An equal number of randomly selected samples from the four academic years was used to allow for variation in terms of the participants' language proficiency level. This was considered an important factor as participants in the ER programme would select and read books at different language levels. However, due to the design of the research, which involved matching participants in the two research groups, different sampling procedures were employed to choose a control partner for each participant in the experimental group. Assigning a control partner to each experimental participant was an essential method, since the 40 experimental participants would be assessed through individualised tests. The aim of the

two groups' involvement therefore, was to observe the effect of their university language course on their learning of the target words, and to ensure that the experimental reading treatment was the only source for incidental vocabulary gain. However, the primary concern was that it could be problematic if the corresponding participants had vastly different repertoires of vocabulary. In order to solve this challenge, the participants' scores from the VLT discussed above were utilised, to determine a match for each corresponding pair in the two groups.

Because the boundary between the two target zones of words was made above the 2000 frequency word band limit, results for all three of the remaining levels in the test (3000, 5000, 10000) were subsumed under one off-list level. The selection of the control partner was made according to the fact that s/he had obtained the same score as the experimental participant for both of the two frequency levels (i.e. the 2000 frequency level (2K) and the off-list level). In cases where exact correspondences were not found, a participant who obtained the closest score (i.e., either +1 or -1) was selected. This formula illustrates the procedure according to which all partners were assigned: Step 1: 2K match + off-list match. Step 2: 2K match + off-list +/-1. Step 3: 2K +/-1 + off-list match. Step 4: 2K +/-1 + off-list +/-1. In all cases the total difference in scores between participant pairs did not exceed four marks (see Appendix 4).

Once the selection of the 40 participants in each group was completed, they were provided with a written research information sheet, and were given sufficient time to read about the research aims, procedures, and their role in the research before being asked to give their consent. Three students (two from the control group and one from the experimental group) withdrew due to family and private circumstances. These three withdrawers were replaced by three other students, following the same random selection procedures mentioned above.

5.2.4 Tasks for the two groups

Mackey and Gass (2005) identify two types of experimental research design: the ‘comparison group design’ and the ‘control group design’. In the comparison group design, the experimental and control participants are offered the same treatment and pre and post-tests. The control group design is similar to the comparison group design, in that, like the experimental group, the control group would take the same pre-and post-tests, but would not receive the same treatment between the two tests. In this research, however, using the comparison group design was not feasible and the control group design was instead adopted for two main reasons. First, because of the free nature of extensive reading, it was not possible to offer a similar and comparable language learning task for the control group participants. In other words, it was not possible to determine the time a control participant’s partner would spend reading in each session. Second, the goal of the two group design in this research was not to compare the acquisition of new vocabulary from two different learning settings (e.g. extensive vs. intensive reading). Instead, the primary aim of involving a control group, was to examine whether other language sources, outside the treatment (e.g., the participants’ traditional language classes), had any effect on any new reported gains in the target words.

The participants in the experimental group were expected to extensively read their selected, graded novels during two-month programme. This group was engaged in both the ER treatment, and four vocabulary tests: a pre-test, first post-test, second post-test and a delayed test. The control group, on the other hand, did not take the treatment, but was involved only in the pre and first post-test. Because of the individualisation of the research readings and testing design, every control participant took the same test as his/her matched experimental partner. Although they were informed of the research into the ER program, their role as a control group in the research was not mentioned. They were not informed about the ER group’s task, nor, did they know that they had

been given the same individualised vocabulary tests as the experimental group. Instead, they were told that they were involved in a study to evaluate the outcomes of their university course. While the experimental group participants were taking part in the ER programme, their corresponding control partners were not restricted to any particular programme or activity.

5.3 Research procedures

This section focuses on the procedures involved in the implementation of the ER programme. It gives an account of the materials selected by participants, and the corpus work involved in determining the target words from their self-selected texts that would be used for testing. The ER library groundwork and the reading treatment is also described.

5.3.1 Participants' choices of books

In order to identify a suitable level of text for each participant in the ER programme, two major steps were applied. The participants were first asked to choose books, from the 40 titles, that they were interested in reading during the eight-week 'treatment'. All titles were presented in three sets according to their level of difficulty. Participants were given two days to check the level which matched their language proficiency. They were told about the level differences, and were asked to read one or two pages from several books at the same graded level. After reading pages from a particular level, the participants were required to report whether the pages they had read were: a) very easy to read; b) difficult and had too many unknown words in the texts; or c) the pages were at their language proficiency level. The book level was decided to be appropriate to the participant's language level if he/she reported "C". Six students chose this statement for books at two levels, but fortunately, none of them omitted to select one of the three levels.

In order to check whether the participants had selected the correct level, they were further asked to sit three level tests (see Appendices 3, 4, and 5) provided by the graded materials' publisher. The three levels tests (stages, 3, 4, and 5 respectively) were administered to the ER group over three days. The participant had to read through a passage and fill in each of 30 gaps with one of 4 possible answers. As discussed earlier in section 4.9.2.1, according to the publisher, the level is considered appropriate for the reader if s/he obtains scores of between 80% and 95% (about 24 to 28 correct out of 30 answers). The tests confirmed the levels already selected by 27 out of the 40 participants. The remaining 13 participants either failed to gain the 80% mark, or scored higher than 96%, so they were reappointed to their appropriate levels on the basis of their performance in the tests. Most of the participants obtained the pass score for one reading level, but there were six participants who similarly gained the appropriate scores for two different stages.

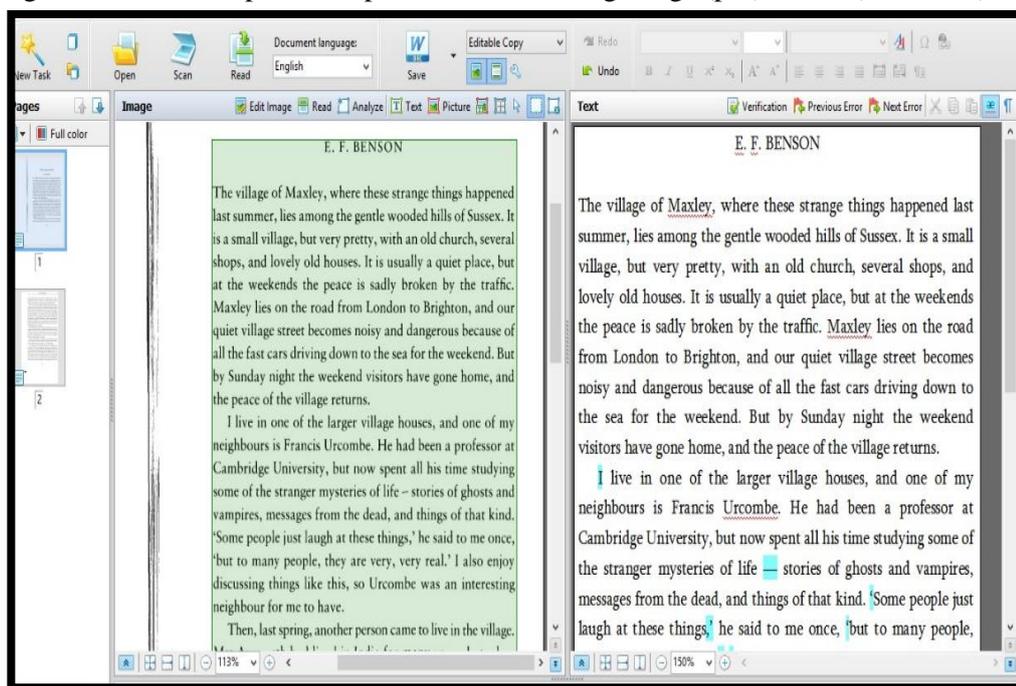
Based on their level of reading, each participant was encouraged to select titles, which s/he was interested in reading during the ER programme. Participants who were found to be eligible for two reading levels were allowed to choose freely and read at one or both levels. By the end of these procedures, the participants had variously selected between four to nine titles to read. In the pilot study, subjects spent a maximum of 22 hours reading a book, so it was assumed to be feasible that the participants in the main study would have enough time to read their selected books. If 10 hours (from 8am to 6pm) was available for participants per university day, then there would be 60 hours in each experimental week (6 days). Since part of this time was devoted to classroom instruction (about 18 hours per week), this left about 40 hours that participants could use for ER every week. This meant that the two month ER programme offered about 320 hours for reading (40 hours \times 8 weeks), which should have allowed a participant to read a total number of about 14 books (total number of experimental hours = $320 \div 22$ hours = an approximate time for reading a book).

5.3.2 Scanning (corpus work)

To determine which were the target words, a method inspired by Horst (2005) was used. The innovative approach used for the pre-post intervention vocabulary measurement in this research was put in place to identify the participants' choice of readers before they began the ER program. In her study, Horst (2005) failed to administer the same pre- and post-test because the participants chose to read different titles. In this research, the very early selection of books allowed for the choosing of the test items and the setting up of individualised pre- and post-tests for the target population. Horst also scanned part of each book to create two different lists of pre-and post-test items. The present study, therefore, aimed to select the individualised target words from all of the texts read by each participant. To be more precise, the pre-and-post-tests included actual words, based on their actual number of occurrences in the books read by the participants during the experiment period. This required a massive amount of groundwork, including book scanning and corpus analysis of the 40 graded reader titles, a process which took nearly 3 months. Four months before the experiment started, all 40 titles were scanned and an electronic folder was created for each title. The optical character recognition (OCR) Abbyy FineReader software was then used, to turn the 40 scanned pdf format texts into editable word documents (see Figure 5-1 below). This particular software was recommended as one of the best available OCR software⁷.

⁷ It has received awards from number of IT organizations, such as the IT publication PC Magazine's "Editor's Choice" award, for its accurate and high quality characters recognition (available at: <http://finereader.abbyy.com/#>).

Figure 5-1: An example of the process of converting image (pdf) to text (document)



Once editable versions of the books had been created, each book was then entered into the VocabProfile software. As mentioned earlier when describing the piloting procedures, this software classifies the words in the text into four frequency bands which appear in different colours (see Figure 5-2). Blue is given to the 1000 most frequent word families (1K); green is for the second most frequent 1000 of English (2K), yellow is given to words from academic word list (AWL) and red is for off-list words (words which are not included in the aforementioned bands and proper nouns). This colouring feature makes it easy for programme users to distinguish between the words in the four frequency categories.

Figure 5-3: VocabProfile analysis (overall output) of the entire text of The Big Sleep

Home > VocabProfilers > English (All-arrow-left to preserve settings) > Output

[EDIT-TO-A-PROFILE SPACE](#)

WEB VP OUTPUT FOR FILE: Untitled

Words reorganized by user as 1k items (proper nouns etc): NONE (total 0 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	556	880	12527	85.56%
Function:	(7713)	(52.68%)
Content:	(4814)	(32.88%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	(3742)	(25.56%)
K2 Words (1001-2000):	209	291	933	6.37%
> Anglo-Sax:	(618)	(4.22%)
1k+2k			...	(91.93%)
AWL Words (academic):	10	11	34	0.23%
> Anglo-Sax:	(5)	(0.03%)
Off-List Words:	<u>2</u>	<u>207</u>	1148	7.84%
	775+?	1387	14642	100%

Current profile	
%	Cumul.
85.56	85.56
6.37	91.93
0.23	92.16
7.84	100.00

Words in text (tokens):	14642
Different words (types):	1387
Type-token ratio:	0.09
Tokens per type:	10.56
Lex density (content words/total)	0.47
<i>Pertaining to onlist only</i>	
Tokens:	13494
Types:	1182
Families:	775
Tokens per family:	17.41
Types per family:	1.53
Anglo-Sax Index:	89.51%
<small>(A-Sax tokens + functors / onlist tokens)</small>	
Greco-Lat/Fr-Cognate Index: (Inverse of above)	10.49%

Table 5-3 below contains similar information for the books used in this study.

⁹ **Word types:** We can count the words in the sentence ‘*It is not easy to say it correctly*’ another way. If we see the word again, we do not count it again. So the sentence of eight tokens consists of seven words or ‘*types*’ (Nation 2001, p. 7).

¹⁰ **Tokens** are the running words or the total number of words in a text (Nation, 2001, p. 7).

Table 5-3: Output of Vocabprofile analysis of entire words in the graded books

No	Book stage	Book title	Tokens	Types	Families	1K%	2K%	AWL%	Off-list%
1	Stage3	A Pair of Ghostly Hands	11234	1145	716	89.7	5.0	0.1	5.2
2		Dancing with Strangers	11573	1450	859	86.9	6.2	0.1	6.9
3		Ethan Frome	10137	1126	690	87.0	6.3	0.1	6.7
4		Goldfish	10130	1026	673	86.3	7.7	0.1	6.0
5		Love Story	8378	1012	644	87.2	5.5	0.3	7.0
6		Martin Luther King	9127	993	627	85.9	4.6	0.6	9.0
7		Rabbit-Proof Fence	9820	1129	701	83.8	7.3	0.2	8.6
8		Skyjack	8202	834	575	85.1	7.4	0.8	6.7
9		The Card	10440	1148	719	85.2	6.4	0.1	8.3
10		The Star Zoo	8632	899	585	88.0	5.1	0.9	6.0
11		The Three Strangers	11423	1203	733	89.4	5.6	0.2	4.7
12		Kidnapped	11780	1162	707	87.7	5.3	0.0	7.0
13		The Crown of Violet	10140	1132	692	85.8	5.8	0.2	8.2
14		The Long White Cloud	10899	1233	760	87.8	5.7	0.1	6.4
15		The Prisoner of Zenda	9811	988	643	87.9	4.5	0.1	7.5
16		The Railway Children	9057	975	637	89.0	5.6	0.1	5.4
17		The Secret Garden	10334	1036	634	88.1	5.0	0.0	6.8
18	Stage4	The Scarlet Letter	16162	1444	846	87.1	7.2	0.1	5.7
19		The Silver Sword	16525	1339	837	86.0	6.7	0.1	7.3
20		The Big Sleep	13494	1182	775	85.6	6.4	0.2	7.8
21		A Morbid Taste for Bones	14999	1384	829	83.4	6.1	1.0	9.5
22		The Whispering Knights	15835	1442	872	86.3	6.6	0.2	7.0
23		Gulliver's Travels	14888	1606	935	89.9	6.3	0.2	3.6
24		Three Men in a Boat	17366	1334	805	89.3	6.5	0.2	4.1
25		A Tale of Two Cities	16761	1252	773	86.0	5.1	0.1	8.8
26		Little Women	14139	1254	777	86.3	5.9	0.1	7.8
27		Lorana Doone	16260	1399	825	89.8	4.8	0.1	5.4
28		Land of my Childhood	16345	1890	1104	87.1	7.0	0.5	5.4
29		Dessert, Mountain, Sea	14253	1422	874	83.5	6.6	0.7	9.2
30		Death of an English Man	15955	1293	780	85.9	6.8	0.7	6.6
31	Stage5	The Accidental Tourist	22061	2057	1176	84.0	7.0	0.3	8.7
32		Brat Farrar	21862	1748	1038	85.8	5.2	0.4	8.6
33		The Dead of Jericho	22371	1677	997	84.6	6.7	0.8	8.0
34		Ghost Stories	19843	1776	1041	86.4	7.8	0.3	5.5
35		Sense and Sensibility	22401	1899	1002	85.6	5.8	0.5	8.2
36		David Copperfield	23760	1779	1020	87.5	6.1	0.3	6.1
37		The Bride Price	18527	1691	979	84.8	6.5	0.3	8.4
38		The Age of Innocence	23243	2081	1113	87.1	5.5	0.3	7.2
39		The Riddle of Sands	21268	1874	1083	85.1	6.4	0.8	7.8
40		Heat and Dust	22261	1693	986	86.4	5.2	0.3	8.1

5.3.3 Target words

Investigating incidental vocabulary growth requires a careful selection of the target words. The design for determining the target words in the present study was based on that of Horst (2005), which emphasized ecological validity over control, which meant presenting genuine words in meaningful reading books. Several studies of incidental vocabulary learning (Hafiz & Tudor, 1990; Cho & Krashen, 1994; Grabe & Stoller, 1997) have been criticised for using words that were likely to be encountered outside the reading programmes. Since the research participants were involved in regular EFL courses, the challenge was to find a method which could eliminate the effect of other sources of input on the learning of the target words. To achieve this goal, off-list words were originally targeted for this research (after removing all proper names from the off-list frequency output). This is because, it is suggested that words in this low frequency category are less likely to be encountered in school textbooks (Horst, 2005). However, using the off-list words in the pilot experiment tests posed two major problems for the main research design. Firstly, the word frequency analysis of the books at the three target levels showed that the texts contained a low proportion of off-list word families, in fact less than 10 % (as illustrated in Table 5-3 above). The vast majority of running words in the texts (over 80%) came from the most frequent 1000 word families (K1), this is because simplified texts are very likely to include fewer off-list words than texts which are originally written for native speakers (Horst, 2005). Given the fact that a word in every 1500 words in the texts was targeted in this research, it was believed that the low proportion of off-list words alone would not cater for the number of items that would need to be included in each individualised test.

To address this test design challenge, it was decided to include words from the 1001-2000 frequency level (2K) amongst the vocabulary test items. Items from the 1000 most frequent zone however, and from the AWL were excluded. This was because of their high frequency of use; many words in the 1000 frequency zone were believed to have already been familiar to the participants, whereas,

as Table 5-3 above indicates, few AWL words were found in each reader (less than 1%). For this reason, all individualised vocabulary tests were comprised of two-thirds of items from the off-list frequency band, and one third of items from the 1001-2000 most frequent word families in English.

Several researchers have used substitute or unreal words when investigating incidental vocabulary acquisition, so as to be sure that learning resulted from reading. (Waring & Takaki, 2003), for instance, replaced 25 words from five repetition groups (i.e. 1, 4-5, 8-10, 13-14, 15-18 groups) in one graded reader they gave to their subjects (*A Little Princess*). Despite the fact that substituting the L2 forms of the target words with nonsense words in the texts could have given an advantage, in terms of the ability to the attribution of any knowledge improvement solely to the treatment, it was not possible to apply such approach in the present study for three main reasons. Firstly, the participants in this study encountered the target words repeatedly in different graded books, whilst previous studies, which used nonsense items, replaced a very small number of real target words in short reading texts. Second, this study measured learning of a much larger number of target words, occurring at extremely high frequencies of repetition. Last but not least, for ethical reasons, it was not possible to expose learners to repeated nonsense words without their knowledge. Therefore, for the aforementioned reasons, it was not practically feasible to use nonsense words in the present study. However, the risk that the target words in the present study could have been picked up outside the treatment was monitored by assessment of the control group's performance before and after the ER treatment.

Although new incidental learning of vocabulary from reading was measured through the actual words encountered by the participants in the reading materials, some fake words were partially involved in the testing procedures. These words looked like plausible English words and took on English spelling conventions. One fake word was tested with every five real words included in

each individualised test. The purpose of including these words was to distract the students' attention from the target words. Using this method also allowed the researcher to check whether the students' answers were based on any guesswork.

The number of words in the experimental books varied according to their stage or level, ranging from approximately 10,000 words in books at stage 3; 16,000 words in books at stage 4; to 23,000 words in books at stage 5. For testing purposes, the number of stories a participant selected was multiplied by the number of words encountered, at the level s/he chose to read from. For example, they were multiplied by 16000 if a participant selected six books from stage 4 ($6 \times 16000 = 96000$). As one word in each 1500 words a participant read in this experiment was targeted, the sum was then divided by 1500 in order to determine the total number target words ($96000 \div 1500 = 64$). For test item inclusion, as discussed above, the target testing number was divided by three ($64 \div 3 = 21.33$). This resulted in one third from the 2K group of words (21.33) and two thirds of off-list items (42.66). Where decimal fractions were involved, low decimal fractions were rounded down (e.g. 21.33 was rounded down to 21), while high decimal fractions were rounded up (e.g. 42.66 was rounded up to 43), resulting in the choice of 21 2K words (roughly one third) and 43 off-list items (two-thirds of the total). A fake word was included in each five target words as a distractor, so, in the example above, 13 non-real words were added to the 64 target items ($64 \div 5 = 12.8$). Here, similar fraction calculation procedures were applied to the entire number of distractors. In all cases where the fraction was above (.05), an additional fake word was added to the whole number. Therefore, the aforementioned example of the individualised test ultimately included 21 2K items, 43 off-list words, and 13 distractors, bringing the total number of test items to 77 words.

Based on the above criteria, very careful procedures were followed to ensure, as far as possible, an equal distribution in terms of word class across the

individualised tests. However, because the target items were selected on the basis of one out of each 1500 words occurring in the books a participant had chosen to read, the number of words included in the individualised vocabulary tests varied, as some participants had read more than others. In other words, the lengthier the texts a participant had read, the more words were included in his/her individualised test. The fact that the participants read a different number of books which varied in terms of level and length made it difficult to create a standardised test applicable to all participants in this study. Taking into account this variation in the participants' individual amounts of reading and of tested words, it was important to adopt a reliable approach when dealing with the individualised acquisition rates analysis in Chapter 7. The aim of the very robust method used was to measure individualised vocabulary gain percentages rather than the number of words learned by each participant. It was assumed that percentages are more valuable for comparing data where the totals are different. This calculation approach provides an opportunity to produce more robust results when reporting the individualised incidental gains later in Chapter 7.

The selection of the target vocabulary for individualised tests from the two frequency categories was based on several criteria. First, members of one word family were treated as a single word. This was based on Bauer and Nation's (1993) guidelines for the Level 3 scale. At this level, a family is considered to consist of a baseword, along with inflected and derived forms that use frequent and regular affixes. For example, *bore*, *bored*, *boring*, *boringly* were treated as four members of one word family. Second, words which had two or more occurrences in stories were included in the test. However, because of the individualized nature of the readings and tests in this study, it was impossible to have same number of repetitions for items in different individualized tests. Therefore, each test consisted of words occurring different number of times in the books each participant selected. In general, the number of repetitions of target words in all tests ranged from between two and 113 times. Words with one incidental encounter were not targeted, as previous research had shown

that a single meeting of new vocabulary seldom resulted in full acquisition of word meaning (Nagy et al., 1985). Moreover, if two words had the same number of repetitions, the priority was given to those which appeared in more than one book. This was to ensure repeated circulation of the target words throughout the programme period. Because of the nature of the vocabulary measurements in this study, it was impossible to systematically control the word class (whether there was an equal distribution of nouns, verbs and adverbs in each individualised test) and word length across the various individualised tests, which contained different number of words.

Overall, from those that met these criteria, 667 words were eventually tested. These words arbitrarily represented the four major word classes (i.e., noun, verb, adjective, and adverb) and were divided into 528 words from the Off-list frequency band (see Appendix 5) and 139 items from the 2K band (see Appendix 6). While the vast majority of these words were tested in a single individualised vocabulary test, some occurred variously in more than one test. Table 5-4 below displays the number of target words from the two frequency bands and their occurrences across the different individualised vocabulary tests.

Table 5-4: The target words and the number of their use across the individualised vocabulary tests

Word category	Total items used in the study	The number of times words used in a single or various individualised tests											The number of times words occurred across the individualised tests
		1	2	3	4	5	6	7	8	9	10	15	
Off-list items	528	353	101	45	20	9	0	0	0	0	0	0	815
2K items	139	55	33	16	8	4	9	4	3	5	1	1	397
Total	667												1212

5.3.4 Library and readers

This ER research, designed for use of Libyan participants, proceeded with a level of care similar to that in the ER programmes described in the literature review chapter. Thus, the treatment design carefully considered the necessary

key features, such as the need for participants to read large amounts of material, representing various topics and genres at their level, as these are essential elements of any successful ER programme (Day & Bamford, 1998). In the preparation phase, a mini-library was created for the reading treatment in one of the English Language Department rooms. Representing various genres, such as classics, crime and mystery, fantasy and horror, true stories, the graded readers used in this study included a variety of highly interesting titles, chosen from the Oxford University 'Bookworms' collection (see Appendix 7 for sample of reading material). Based on the results obtained from the pilot study, three sets of books at the three different stages were provided, making a total of 40 titles (i.e. 17 titles from *stage 3*; 13 titles from *stage 4*; 10 titles from *stage 5*). Such variation in terms of language levels and genres is a crucial principle for any ER programme (Day & Bamford, 1998, 2002), because it gives the participants more opportunities to select texts which cater for the assorted nature of their interests and the potential discrepancy in their language proficiency. More titles from the lower stages were offered because the word count in a book at Stage 5 (23.000 words), for example, was about double of that at Stage 3 (10.000 words). Table 5-3, presented earlier, demonstrates the number of books belonging to each target book stage that were chosen. It also shows the variation, in terms of the running number of words in the texts, at the three levels.

Three copies of each of the 40 titles were purchased, totalling 120 books. Providing additional copies aimed to remove restrictions on what a participant wanted to read first. In other words, by having three copies of each book, it was more likely that the titles were always available and participants could read their selected stories in their preferred order. For the participants' convenience, the books were classified according to their levels and kept in a small bookcase. They were organised using the publisher's coloured labels. For example, the red colour referred to books in stage three, the purple label was attached to stage four books, and the light blue referred to stage five books. This classification was established to help the research participants to easily

borrow and return their selected novels. In order to simulate, as much as possible, the recommended conditions for an ER programme (Day & Bamford, 1998), the library was provided with a comfortable sofa, cushions, chairs and a long table. Likewise, free of charge refreshments were offered during the reading time. All these facilities were added in order to increase the relaxation of the participants and therefore encourage their reading.

5.3.5 The treatment

Prior to the treatment commencing, considerable groundwork and planning were implemented, to ensure that all the experimental participants were engaged in the reading programme. Similar to other Arab EFL contexts (Al-Homoud, 2007), it was believed that ER was a novel approach in Libya and most, if not all, of the students in the English Language Department considered reading in English to comprise what is called ‘intensive’ reading. Another concern was that the target participants did not have a reading habit. They did not have the experience of reading longer passages in English. This is because, for many Libyan EFL learners, reading is considered as one of the school curriculum requirements, which they practise to pass their examinations, rather than as a source of pleasure.

Therefore, before starting their reading, the intention was to foster the 40 experimental participants’ motivation, so they would be encouraged to read, as far as possible, all their self-selected titles. Participants were told that reading their chosen texts could enhance their English language learning. The participants were not told that they would be tested after the treatment, as that could invoke the process of intentional learning of vocabulary. Additionally, an explanation of the library use and book checking out system was given. They were informed that books would be kept in the library and they were not permitted to be taken away. In addition, participants were not allowed to use dictionaries. Dictionary use was not recommended as incidental growth could not be exclusively ascribed to the reading treatment when dictionary use was

involved, a fact which has led to the findings in a number of previous studies (e.g., Cho & Krashen, 1994; Grabe & Stoller, 1997) being treated with caution. More importantly, frequent looking up of word meanings from a dictionary is believed to interfere with short term memory and, as a result, interrupts the process of reading comprehension (Knight, 1994). With these primary concerns, therefore, the participants were encouraged to use the surrounding context to guess the meaning of unfamiliar words. They were also recommended to skip the word if they could not guess its meaning.

To make the library accessible to all participants, two of the department librarians were involved in administering book lending and library management. Once the reading programme began, the library was open six days a week (Saturday to Thursday). The participants were free to carry out daily reading, during their regular university hours. The time each of the participants spent on their reading sessions varied depending on the participants' free time and their class timetable. Records, for instance, showed that some participants had a single very short reading session whereas others had more than one visit to the library in one day. Similarly, some appeared each day of the experimental session, while in other instances some had a number of absences.

A filing system was used to document the programme and record both the titles read and the participants' instant reports about the amount of actual reading, comprehension, and difficulties they had encountered (see Appendix 2). Forty folders, with printed labels, bearing each of the participants' name were kept in the library. A list of titles, chosen by the participant, was kept in his/her folder. During the experimental phase, participants were required to keep a record of time and the number of pages read in each of the reading sessions, for each book. The time record form (see Appendix 8) had to be folded in half and kept inside the book. The aim here was threefold; first, it allowed the researcher to track the participants' reading; second, having the time record form left in the

book served as a sign that the book was in use and should not be taken by other participants. Moreover, the participant would know the page on which s/he stopped in her/his previous reading session. Every participant was asked to tick off a book in the list once s/he had finished it. Following that, the participant was asked to fill in the instant book report form and keep it in their folder. For the purpose of monitoring the participants reading during the programme, the report was important, as it allowed the researcher to check whether the participants had really done their reading. Additionally, it aimed at gathering data related to the participants' reading interests, comprehension, and difficulties they may have had encountered when reading their books.

During the treatment, the researcher kept checking the programme procedures with the librarians. Each participant's record file was regularly checked to track his/her reading progress. The validity of the participants' instant book reports on their readings was also checked through the opportunity sampling approach, which, as Wellington (2000) states, is used to choose samples from people who are available at the time of the study. In total, the researcher checked seven reports during the two-month reading programme. The selection of the report was made on the basis of book completion at the time of the report being checked. The researcher sought to check participant information after s/he had finished one of their selected books. Information about the participants' reading progress, comprehension, and other issues such as reporting and filing were elicited. All seven reports indicated that participants were positively involved in the reading programme, so there was no serious concern in relation to the materials or the participants' reading enjoyment and comprehension.

After the six weeks of reading, some of the research participants began to withdraw from the study. By week eight, about half of the participants could not continue the programme due to the unforeseen circumstances in Libya, i.e., civil war, during the time the research was conducted. Since they withdrew at a very advanced stage of the research period, it was inappropriate to replace

them with other students. Out of the 40 experimental participants, only 18 managed to complete the program. In addition, the ER program was originally planned for eight weeks. However, at the end of the programme, some of the participants still had not finished reading all their selected books. Therefore, to allow more time for these participants to finish the books which were still being read at that time, the programme was extended for another week. At the end of week 9, 18 participants were reported to have had successfully finished the books, and only two of them did not manage to read all books on their selected list. Thus, they were advised to continue their reading in the following week (week 10) where they spent between 2 to 3 days to finish books in hand. At this point, one book in the lists of these two participants was found to be unread, so words that appeared solely in these two books were excluded from the two participants' tests. Similarly, if a word appeared in other books these two participants selected, then the number of word occurrences in the two unread books was eliminated from the total number. The records showed that a total of 117 books were read by all experimental group participants during the treatment. These books represented 39 different titles, as there was one book, '*A Tale of Two Cities*', that none of the participants had been interested in reading. As shown in Table 5-5 below, almost all the titles provided in the ER programme were circulated, and the number of times a title was read by different participants during the treatment varied between one and six times.

Table 5-5: The number of times each of the experimental books circulated around the participants

No	Levels	Titles	Participants																No of times a book circulated			
			P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16		P17	P18	
1	Stage 3 books	A Pair of Ghostly Hands				√	√		√												3	
2		Goldfish			√	√	√															3
3		Love Story	√	√			√															3
4		Martin Luther King		√																		1
5		The Crown of Violet		√	√	√	√															4
6		Ethan Frome			√		√		√													3
7		Rabbit-Proof Fence		√	√																	2
8		Skyjack		√																		1
9		Dancing with Strangers			√	√	√															3
10		The Card	√			√				√												3
11		The Star Zoo	√	√		√																3
12		The Three Strangers	√			√				√												3
13		Kidnapped		√	√		√	√														4
14		The Long White Cloud	√			√	√	√	√													5
15		The Railway Children	√						√													2
16		The Prisoner of Zenda			√		√															2
17		The Secret Garden	√	√		√		√	√													5
18	Stage 4 books	The Scarlet Letter								√	√	√									3	
19		The Silver Sword								√												1
20		The Big Sleep										√										1
21		A Morbid Taste for Bones								√	√		√	√								4
22		The Whispering Knights									√	√	√								√	4
23		Gulliver's Travels									√		√	√							√	4
24		Three Men in a Boat								√	√			√								3
25		A Tale of Two Cities																				0
26		Little Women								√		√										2
27		Lorana Doone								√	√											2
28		Land of my Childhood										√										1
29		Dessert, Mountain, Sea										√		√							√	3
30		Death of an English Man									√	√		√	√							4
31	Stage 5 books	The Accidental Tourist															√	√	√		3	
32		The Age of Innocence												√	√	√	√					4
33		Brat Farrar												√		√						2
34		The Dead of Jericho												√	√	√		√	√			5
35		Ghost Stories											√		√	√		√	√	√		6
36		Sense and Sensibility												√	√		√					3
37		Heat and Dust													√		√					2
38		David Copperfield															√	√	√			3
39		The Bride Price															√		√			2
40		The Riddle of Sands											√		√	√	√	√				5

5.4 Research instruments for data collection

A variety of research instruments were used in the process of gathering data for the present study. Every research instrument is particularly suitable for certain sources of data, yielding information of the kind and in the form that can be most effectively used. The data sources for this study were varied and were designed to address the range of research questions posed. The following is a description of each data source.

5.4.1 The vocabulary test

In this study, the vocabulary knowledge scale (VKS) was used to assess incidental vocabulary learning from an ER programme. A scale, developed by Schmitt and Zimmerman (2002), was adopted to measure changes in the participants' receptive and initial productive knowledge of target words after they had finished reading. As previously mentioned, one major concern, which was noticed in the test, was that the demonstration of knowledge for stage "C" (I understand the word when I see it in a sentence, but I don't know how to use it in my own writing) and "D" (I know this word and can use it in my own writing) was based on the participant's judgment and I was not fully confident that the subject's statement would really reflect his/her knowledge of the word. To improve the test reliability, the scale was modified by asking the participants to provide an Arabic translation, along with these two choices.

5.4.1.1 Test administration

This section discusses the four-time administrations of individualised vocabulary tests: pre-test, 1st post-test, 2nd post-test, and delayed test).

5.4.1.1.1 Pre-test

One of the limitations that some of the previous reading research has suffered from is that the measure of incidental word gain was based on the subjects' performance on vocabulary tests after they had finished the reading. However,

the extent to which research subjects were familiar with the target words before the reading treatment was not examined. Even when this was assessed, the instruments were not sensitive enough to show partial knowledge of the target words. In these cases, there was some ambiguity about the number of words that were really obtained as result of incidental encounters through reading. For example, it is difficult to accept a 20% gain of 30 French words after just 40 minutes reading, as reported by Dupuy and Krashen (1993), without questioning how many of these words had been unknown to the subjects before they started reading. Using different target items for the pre and post-tests, Horst (2005) investigated incidental gains from two frequency bands, but eventually she could only argue for the results of very low frequency words (off-list), as other sources could have affected the learning of words from higher frequency levels of use (2K). Indeed, the current study tested words from the same two bands of frequency, but in order to avoid this limitation careful control over sources of learning was established.

What made the current study unique, were the procedures incorporated in the design of how the target words were chosen, and the testing, as was discussed earlier in sections 5.3.2 and 5.3.3. Although similar procedures were applied by Horst (2005), she was not able to administer the same items in the pre-test for her subjects as in the post-test given after they had finished reading. However, the advanced self-selection of the graded materials in this study, allowed the researcher to target words from the books each participant in the ER group had selected to read during the programme, in order to address this challenge.

The use of authentic L2 words in studies investigating incidental acquisition from L2 reading has been one of the strongest challenges for researchers, and one that they have largely failed to deal with. Many of the studies discussed in Chapter 4, used modified “nonsense” words to avoid any possibility of the words being already known by subjects. To the research’s knowledge, Horst (2005) was the only researcher who created a corpus from the reading materials

in order to select the test items from the books read by her subjects. However, her corpus was created from only twenty pages of each book used in her research. Horst (2005) acknowledged that such an incomplete corpus of graded readers could not provide an adequately clear picture about the learning opportunities for words encountered in all of the texts read by the participants. She concluded that “[u]ntil it is completed, no definitive conclusions about word learning can be drawn using the proposed methodology” (p. 377). Therefore, in the present research a much larger computer-corpus was built, based on the entire number of pages of all the graded books available. This, of course, allowed for early determination of the target words on which each participant would be tested before they started their reading. Therefore, for the aforementioned reasons, the findings obtained from this study can be argued to be more reliable than those presented in previous reading studies.

At the beginning of the present research, individualised vocabulary pre-tests were administered to the 40 participants in both the experimental and the control groups. The test assessed the participants’ knowledge of the target words for both groups before the ER programme was implemented. The participants in both the experimental and control groups had the test one week before the reading treatment commenced. Each set of matched pairs across the two groups received the same individualised items in the pre-test. A week’s interval between the test and the reading was proposed, in order to distract the participants’ attention from the research aim, and to eliminate the target words from the experimental participants’ minds. The two groups took the test together in two university rooms in the following manner. Using their research number, 20 participants from each group were randomly mixed to take the test in one room. It was assumed that, such a step would help to reduce the risk of the ER group establishing a correlation between the vocabulary test and the reading programme starting a week later. For the same reason, the test was administered by two colleagues who verbally explained how the test should be answered. The participants in both rooms sat the test at the same time. They were told that the test was for the purpose of evaluating their university

courses' evaluation, but they did not know that they were taking the same test as their matched pairs, nor were they informed about their role in the study.

The number of words in each individualized test varied according to the number and the level of the books the experiment participants chose to read. Early estimates from the pilot study suggested that participants would need more or less an hour to complete 100 words in the scale. Considering the lengthy tests for some of the participants, it was necessary to ensure that such an answering process would not be too painstaking, so that there would not be negative effects concerning the test validity as a result of subject being overburdened or fatigued. Therefore, the number of items in each individualised test were divided into lists of a dozen words. The 12 lexical items in each cluster were randomly selected from the two target frequency categories (2K and off-list). Once a participant had finished the first cluster, the next list was immediately given to him/her. Participants were freely allowed to take a rest if they felt tired, providing that the cluster of test items at hand had been completed. This procedure aimed at reducing the possibility of checking the meaning of unfamiliar words if participants decided to take a break, as they would not have been exposed to the rest of the target words.

With regard to the time spent answering the individualized tests, it was found that the participants who selected their readers from the higher stages required more time to answer all items in their tests. The majority of the participants were able to finish the test in one sitting. There were occasions where 11 participants asked for a break and one asked to complete her test the next day. Because of the relaxation strategies put in place, the participants seemed undaunted by the relatively long lists of lexical items. All participants in both groups successfully completed the tests within a maximum of about 70 minutes. After they had taken the pre-test, participants in the two groups were not told about the aim of the test, nor were they told that they would be re-tested on the same words after the reading program.

5.4.1.1.2 First-post-test (one week after reading)

It should be mentioned here that only 36 participants, equally representing the two groups took the post-test, as 44 of the experiment and control participants were excluded along with their matched partners after they withdrew from the study. The 18 participants who had finished the ER programme took the first post-test. With a change in word order, the same individualised pre-tests were administered to both the experimental and control participants one week after the ER program. Having the post-test immediately after the reading program could have posed the risk that the target words were still fresh in their minds, but the seven days' interval reduced this possibility. The two groups had the same unannounced individualised post-tests and the same procedure described in the pre-test was followed in this test. There were 18 participants in each room representing the two groups equally. The researcher and a colleague conducted the test, with a prior instruction on how the test was to be answered given to the participants. Again, there was no indication of the actual aim of the test. Participants were free to take rests if they felt it necessary, on condition that they finished the sheet in hand. As with the pre-test, the time taken to complete the post-test varied, but all participants were able to finish their tests in one sitting, and in 55 minutes maximum.

5.4.1.1.3 Second post-test (two weeks later)

The control group was also involved in the pre-and post- tests to check whether the target words could have been acquired from sources other than the reading treatment. As this matter had already been checked in the first post-test, it was not thought necessary to confirm this again after two weeks. Therefore, only the 18 experimental participants were asked to take the second post-test, which took place in one room with all needed instructions provided. In order to track vocabulary retention over time, the delay between the first post-test and this test was initially one month. However, due to the circumstances in Libya, the research implementation timetable underwent considerable amendments. Because of the delay in commencing the reading treatment, it ended relatively

shortly before the participants had their spring semester final exams (less than one month). Therefore, it was not guaranteed that the participants would be available if the test was delayed by one month. The 18 experimental group participants had the same pre-test, but all individualised tests were shuffled and items were randomly reordered. This would have served to avoid the participants' direct attention to the target words. All the participants finished the test in one day and six of them asked for a break at different points of their tests. All participants were able to finish their tests within an hour.

5.4.1.1.4 The delayed-test (nine months later)

As the study was aimed at exploring incidental word retention over time, the participants in the experimental group were also given a delayed test nine months after they had finished their ER program. Most of the previous research on incidental vocabulary learning from reading gives a "snapshot" of vocabulary knowledge through the immediate post-test. Nevertheless, it has been emphasized that attrition occurs in any learning. Thus, a delayed post-test would be highly informative about the actual gain (Schmitt, 2010). Few studies have tested the retention of incidental word learning over time. To the researcher's knowledge, the lengthiest period of post-test delay to date was conducted by Waring and Takaki (2003), who post-tested their subjects three months after they had done their reading.

In this study, one of the researcher's colleagues administered the delayed test to the 18 experiment participants. The test was taken on different days in one of the university rooms. This was because, at the time of the test, there were four participants who had already finished their final semester subjects and were working on graduation projects. They were not required to attend classes but they had to come occasionally for supervision tutorials. Therefore, on the first day, the test was given to 14 participants who were attending full course subjects, while the remaining four had the test after six days.

5.4.1.1.5 Scoring criteria

The instrument used in this research was a developmental scale developed by Schmitt and Zimmerman (2002). The scale consists of four levels worded this way:

- A) I don't know the word (*no knowledge*)
- B) I have seen the word before but am not sure of the meaning (*slight knowledge*)
- C) I understand the word when I see it or hear it in a sentence, but I don't know how to use it in my own writing (*receptive knowledge of meaning and form*)
- D) I know this word and can use it in my own writing (*productive Knowledge of use in writing*)

For data analysis purposes, several steps were taken to ensure that the test produced accurate results. The tests were scored in the following manner: 0 point was given to words rated at Level A. Words rated at Level B were given 1 point; words were given 2 points if they were ranked at Level C and words rated at level D were given 3 points. Therefore, the type of word knowledge ranged from 0-3. (0= no knowledge, 1= slight knowledge, 2= receptive knowledge, 3= productive knowledge).

A word was marked "D" if a participant was able to provide an equivalent Arabic translation and could use this word appropriately in a sentence. Like Schmitt and Zimmerman (2002), words rated at level "D" were judged according to a basic level of use rather than their complete productive mastery. Thus, the participants' ability to use these words in sentences did not consider the appropriateness of grammatical, stylistic, and collocational aspects. However, for a word to be correctly levelled at "D", the participant should have been able to productively provide a sentence which clearly illustrated the meaning of the word.

In this study, a slight, but important, alteration was made to Schmitt and Zimmerman's scale. The participants were required to provide an Arabic translation when they decided the word was at "C" or "D" level. Accordingly, words which the participants rated as D, but ambiguously used in sentences (e.g., I like monk), were marked at a "C" level (2 points). This is because the participant did not clearly illustrate the meaning in a sentence, but could provide an equivalent Arabic translation, which was required at "C" level. For the sake of reliability, the researcher and a colleague scored each response rated at Level D. The appropriateness of meaning was judged by an inter-rater agreement of 100%. The same step was taken if the participant provided the translation but s/he failed to use the word in a sentence. A word was, however, downgraded to "A" level (0 point) if it was rated "C" or "D", but no other requirements (translation or translation and a sentence) were demonstrated, as it was not possible to verify whether the answer was based on knowledge from any previous encounter of that word, or whether the word was mistaken.

In their study, Schmitt and Zimmerman eliminated from the instrument any subject who rated a non-word at C or D level. However, it was not possible to exclude any of the research participants from the current study, as their number was, relatively speaking, small. Moreover, it was believed that the prior instruction given to participants about the inclusion of the distractor words and the requirement of the Arabic translation (this was not included in Schmitt and Zimmerman's research instrument) at these two levels would prevent any chance of guessing a fake word's meaning. Results indicated that the vast majority of words in this category were rated either at A or B levels. Only a small number of words (9/723) were rated at "C" and "D" (1.2%) by the experimental and control groups in all administered tests. In these cases, it was assumed that participants could have confused their meanings with other words, rather than relying on guessing. Since the distractors adhered to English spelling conventions, it was not surprising that some participants might have believed they had seen them before. Thus, words rated at B level were logically accepted.

5.4.1.1.6 The reliability of self-reporting tests

Whilst participants' self-report of new vocabulary learning has been widely used in other studies (Schmitt & Zimmerman, 2002), it is generally recognised that it is not an entirely reliable measure. This is particularly true when the phenomenon being investigated is a sensitive issue, as is the case with incidental vocabulary acquisition. However, the use of scale level tests, which required the participant to determine one of the four levels at which s/he recognised the word appeared to be very reliable in this study. With the slight alterations made to the original version by Schmitt and Zimmerman (2002), the participants were required to support their chosen answer, particularly at their receptive and productive levels, with an equivalent Arabic translation. Moreover, the inclusion of the fake distractor words in the individualised tests further assessed the reliability of the participants' responses. The results indicate that very few of these fake words were rated by the participants at receptive and productive levels (1.2%). This very small rate was logically acceptable, as some participants could have confused fake items with other English words they might have had encountered previously.

5.4.2 Semi-structured interviews

Cohen et al. (2007) describe an interview as a verbal communication between the interviewer and the interviewee, used to gain important information about a particular topic. Interviewing, as a method for collecting research data, has been very common in educational research. There are three common forms of interviews: the structured interview, the semi-structured interview, and the unstructured interview. The distinction between these types of interviews is made in terms of degree of the flexibility and the type of questions being asked by the interviewer. While the questions of a structured interview are standardised, pre-planned, and in a given order, semi-structured interviews are more flexible and allow new questions and ideas to be generated from the interviewee's responses. Unstructured interviews, however, use more flexible questions that are likely to be subject to constant changes. They run through

free flowing conversation, during which the researcher asks questions without structured guidelines.

With the intention of probing to understand the participants' perspective on the ER approach, I carefully constructed some semi-structured interview questions, based on my reading of the literature. Cargan (2007) argues that questions in the semi-structured interviews are characterised by flexibility and the interviewer may change the order, delete, explain, or even modify questions if s/he believes it is appropriate and facilitates data elicitation. The interviews were one-to-one and were held with 14 of the 18 participants as 4 did not agree to be interviewed. In this study, individual interviews were preferred over a focus group method. This was because individual interviews tend to be more effective "when we may expect a variety of different stories to be told concerning a setting or context [or] where the topic to be discussed is sensitive" (Cohen et al., 2007, p. 8). The fact that research question 4 aimed to explore the differences in the ways participants picked up vocabulary from reading various self-selected and individualised materials, as well as the sensitive nature of investigating incidental learning of vocabulary from reading, made the individual interviews a practical method in this research.

The interviews combined both close and open-ended questions (see Table 5-6), which sought to elicit information related to the participants' individualised reading and their perception of the ER programme. While close-ended questions involve explicit options for respondents to choose from, open-ended questions allow more freedom for respondents to express themselves and solicit further information from them. With the purpose of reducing the possibility of confusion, the interview questions were carefully phrased, using natural and simple language so as to avoid obscure questions or equivocal words.

Table 5-6: Guiding questions and prompts used in participants' interviews

Reading experience	<ul style="list-style-type: none"> • Do you like to read? • What do you like to read? • When and where do you read? • How often and what kind of materials do you read? • Do you read extensively in English in your free time? • Did you enjoy reading the books you selected for this ER programme? • Why did you choose these particular titles? • How did you find the language suitable for your level of English proficiency? • Can you tell me of most interesting things in these books?
Difficulties encountered	<ul style="list-style-type: none"> • Did you find it easy to access the library when you wanted to read your selected books? • What concerns you most about the programme? • At any point of the programme, did you encounter any difficulties in reading? • Did you encounter many unknown words in the texts? • You were not allowed to use the dictionary while reading. Could you tell me how you dealt with the unknown words you met in the books?
Perception about the ER programme	<ul style="list-style-type: none"> • What is your personal opinion about the extensive reading programme? • What did you like/dislike most about the ER programme? • Do you think reading these books helped you to improve your English learning? If yes, how?

The follow-up face-to-face interviews were held three days after the participants had taken the first vocabulary post-tests. Participants who agreed to be interviewed after taking the post-test (14 participants) were informed that the interviews were to collect information about their experience of participation in the ER programme. The interviews were held in the participants' first language, Arabic. This aimed at removing any difficulties interviewees may have faced when attempting to express themselves in English. All the interviews were audio-recorded but only the significant extracts were then translated and transcribed.

So far, a discussion on the research design and procedures employed in preparation of the materials and the tests for this study has been offered in this chapter. This included a detailed description of the process by which this experimental research was carried out and how the data was collected. In the following two chapters (chapter 6 and 7), the data obtained by the aforementioned methods will be analysed and discussed. The results, as well as

the statistical procedures applied to explain the resulting data, will be described. The analysis of the data in the following two chapters will follow the order of the research questions. While chapter 6 will describe findings related to the overall results (answers to research questions 1 to 3), chapter 7 will focus on the analysis and discussion of findings about the individualised reading and incidental vocabulary learning by these participants (answers to research question 4).

Chapter 6:

Overall results

Chapter 6: Overall results

6.1 Introduction

This chapter draws upon the data collected from the individualised vocabulary tests administered four times to the 18 experimental participants and two times to their 18 matched control partners. The chapter presents the analysis and interpretation of the data, and aims to answer three of the four research questions:

- 1) Does reading extensively enhance adult, Libyan, EFL learners' vocabulary knowledge?

To what extent can a two-month ER programme enhance incidental learning of English vocabulary items by Libyan EFL learners for receptive and productive use?

- 2) Does word repetition affect receptive and productive incidental vocabulary learning from extensive reading? If so, how many exposures are needed for incidental vocabulary learning?
- 3) How much of the newly learnt vocabulary is retained 9 months after the ER programme?

Data related to these research questions will be presented in turn. Research question one, is related to data gathered from both the experimental and control groups. Research questions 2 and 3 are related to data gathered from the experimental group.

6.2 Data presentation, analysis, and interpretation

A word knowledge scale created by Schmitt and Zimmerman (2002) was adapted as a testing instrument for this study. Throughout this analysis, the four scale items (Level A, Level B, Level C, Level D) will be respectively

referred to as score marks (0, 1, 2, and 3). Two levels of word knowledge (C, D, i.e., scores 2 & 3) were the main concern here because they demonstrated the learners' receptive and productive knowledge of a word. Although one point was given to a word if a participant thought s/he had “*seen the word before, but was not sure of its meaning*” (score 1), it was not possible to consider word knowledge at this level as “learning” since it did not indicate any type of receptive or productive acquisition. However, this will be treated as ‘*slight knowledge*’ because it indicates an initial stage of word recognition.

For the statistical analysis required to answer the four research questions, I used the SPSS (Statistical Package for the Social Sciences) programme (version 18.0) to analyse the numerical data which was collected through the vocabulary tests. Prior to entry into the SPSS, this data was coded and carefully checked for errors. The results for the research questions are presented in order below. Data presentation for each question will be followed by a summary of the main findings.

6.3 Research Question 1: Does reading extensively enhance adult, Libyan, EFL learners' vocabulary knowledge?

To what extent can a two-month ER programme enhance incidental learning of English vocabulary items by Libyan EFL learners for receptive and productive use?

6.3.1 Experimental vs. control groups' prior knowledge of target words

According to Schmitt (2010), the main benefit of using a pre-test in vocabulary development research is to ensure that there is no pre-existing target word knowledge, which can jeopardise the vocabulary gain results. A number of previously mentioned studies have been criticised for building their findings on the post-test alone. For example, it is difficult to accept Dupuy and Krashen's

(1993) 20% gain of 30 French words, after just 40 minutes of reading, without questioning how many of these words had been unknown to the subjects before they started reading. Using different target items for the pre and post-tests, Horst (2005) investigated incidental gains from two frequency bands, but in the end she could only argue for the results for the very low frequency words (off-list), as it was unclear whether other sources could have affected word learning from the higher frequency level of use (2K). The current study tested lexical items from the same two bands of frequency, but careful control over the sources of learning was established in order to avoid such issues.

As stated in the methodology chapter, each participant in the control group was matched with an experimental participant, according to their performance in the Vocabulary Level Test. The challenge of distinguishing the learning source of the target words was addressed by monitoring the performance of the control group in both the pre-and post-tests. Before measuring the participants' acquisition of the target words, it was necessary to check whether my data allowed for *parametric* or *non-parametric* statistics¹¹. As Table 6-1 and Table 6-6 (below) show, results obtained from the Kolmogrov-Smirnov test¹² confirmed that the data satisfied the assumption of normality (the sample distribution is normal) and therefore a parametric statistical approach was feasible for the statistical analysis of the data.

Schmitt (2010) cautioned against the use of parametric statistical analysis of data obtained through research using word knowledge scales by saying:

The analyses are almost always inferential (t-test, ANOVA, correlation, etc), which require the use of an interval scale, where the distance between the intervals is equivalent and

¹¹ Parametric and nonparametric are two broad classifications of statistical procedures. Parametric methods are used when we know that the population is approximately normal, whereas nonparametric methods are used when the population is not normally distributed.

¹² Kolmogrov-Smirnov test quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution

consistent. With the [vocabulary knowledge] scales currently available, this assumption cannot be met (p. 218).

Taking this caution into consideration, an equivalent non-parametric test (Wilcoxon Signed Ranks Test) to the parametric t-test was also run with the same question 1 research data. In fact, comparison between the results obtained from the two statistical analysis approaches showed that the difference was only fractional, which did not lead to significant difference in terms of p value. Therefore, I decided to follow the parametric paradigm to answer research question 1. The parametric t-test statistical data are presented below in Table 6-2, Table 6-3, Table 6-4, Table 6-7, Table 6-8, Table 6-9, Table 6-10, Table 6-12, Table 6-13, Table 6-15, and Table 6-16. The non-parametric Wilcoxon Signed Ranks Test output equivalents for results in these tables also provided in Appendix 9.

Table 6-1: Test of Normality

Word Category	Group	Kolmogrov-Smirnov test	
		Statistic	p-value
<i>Off-list</i>	Experimental	0.153	0.200
	Control	0.200	0.056
<i>2K</i>	Experimental	0.138	0.200
	Control	0.125	0.200
Total	Experimental	0.155	0.200
	Control	0.182	0.118

Table 6-1 above displays the tests for normality of the data using the Kolmogrov-Smirnov test. At the 5% level of significance, there was no evidence to reject the null hypothesis (the null hypothesis states that there is no difference between two sets of data) that the data was normal. Hence, to compare whether there was a significant difference between the experimental and control groups' mean prior knowledge, a parametric Paired-Samples t-test was performed. The Paired Samples t-test compares the means of two variables. It computes the difference between the two variables for each case, and tests to see if the average difference is significantly different from zero.

The participants' average score for recognition of the target words, as well as the mean percentages of words rated by participants at each of the four knowledge levels in the scale, were calculated and used in this comparison.

The assumption was that words belonging to the off-list zone were considered unlikely to be known by the Libyan EFL learners and that 2K items were more likely to be familiar to them. However, as Schmitt (2010) cautions, researchers should be aware of the fact that learners do not learn new vocabulary in strict frequency order, and they may acquire some low frequency words before learning the most frequent ones. Thus, it was prudent to take a closer look at the participants' prior knowledge of items from each of the two target frequency ranges, and to examine whether the careful procedures for selecting high frequency words from the 2K range had proved effective in this study. The analysis of the overall scores in the two categories is presented first, and then words from each frequency band are individually examined.

6.3.1.1 Overall pre-existing target word knowledge

The descriptive statistics (mean, standard deviations) of the test scores gained by the two groups of participants are presented in Table 6-2.

Table 6-2: Comparison of the experimental and control groups for the overall pre-vocabulary test scores assessed by means of the Paired-Samples t-Test

Group	N	Mean	SD	<i>t</i>	Sig (2-tailed)
<i>Experimental</i>	18	0.52	0.14	1.42	0.173
<i>Control</i>	18	0.47	0.13		

Mean score= 0-3

As the output of the Paired-Samples t-test shows, the experimental group's mean score on all the target words in the pre-test was slightly higher (M= 0.52, SD = 0.14) than the control group's overall mean score (M= 0.47, SD= 0.13).

However, statistically, the difference between the two groups' means was not significant, as the p value was greater than 0.05. These data suggest that, overall, participants in the two groups had the same level of target word knowledge before the experimental group started the reading treatment.

To look at the difference between the mean scores achieved by the two groups of participants for each frequency band, a Paired-Samples t-test was performed independently for the off-list and 2K items; the outcomes are shown respectively in in Table 6-3 and Table 6-4 below.

6.3.1.2 Pre-existing off-list word knowledge

Table 6-3: Comparison of the experimental and control groups on pre-test off-list item scores, assessed by means of the Paired-Samples t-test

Group	N	Mean	SD	<i>t</i>	Sig (2-tailed)
<i>Experimental</i>	18	0.42	0.10	0.18	0.184
<i>Control</i>	18	0.37	0.15		

Mean score= 0-3

6.3.1.3 Pre-existing 2K word knowledge

Table 6-4: Comparison of the experimental and control groups on pre-test 2K items scores, assessed by means of the Paired-Samples t-test

Group	N	Mean	SD	<i>t</i>	Sig (2-tailed)
<i>Experimental</i>	18	0.74	0.31	0.69	0.500
<i>Control</i>	18	0.69	0.21		

Mean score= 0-3

Similar to the overall results in Table 6-2, the off-list and 2K word knowledge data in Table 6-3 and Table 6-4 also show that there was no significant difference between the experimental and the control group in terms of their performance in the testing of items from the target categories. Although the

mean average scores for the off-list (for the experimental group $M= 0.42$, $SD= 0.10$, and $M= 0.37$, $SD= 0.15$ for control group) and 2K words (for the experimental group $M= 0.74$, $SD=.31$, and $M= .69$, $SD= 0.21$ for the control group) were slightly higher for the experimental group, the difference between the two groups was not statistically significant. Since the p-value in the two categories of the target words is greater than 0.05, at 5% level of significance, it is reasonable to conclude that the mean scores of the experimental and the control group can be considered the same. This result implies that the experimental and the control group had largely the same level of knowledge of the target lexical items when they were pre-tested.

To gain an indication of the extent of the participants' pre-existing knowledge in each of the two word frequency categories (Off-list + 2K), the number and percentages of words rated by the two groups at each level of knowledge in the pre-test were calculated and are presented in Table 6-5 below.

Table 6-5: Percentages of words rated at the each level of the scale in the pre-test by the experimental and the control groups

Knowledge Level	Group	<u>Off-List</u>		<u>2K</u>		<u>Overall</u>	
		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
0* (no knowledge)	Experimental	552	67.7	190	47.9	741	61.1
	Control	594	72.9	210	52.9	804	66.3
1 (slight knowledge)	Experimental	195	23.9	140	35.3	335	27.6
	Control	150	18.4	119	30.0	269	22.2
2 (receptive knowledge)	Experimental	47	5.8	32	8.1	79	6.5
	Control	45	5.5	46	11.6	91	7.5
3 (productive knowledge)	Experimental	21	2.6	35	8.8	56	4.6
	Control	26	3.2	22	5.5	49	4.0

*=score mark

N= Number of words according to the four knowledge score marks

%= out of total in each word category

The table displays the number and percentages of the target words rated by the two groups of participants at each of the four levels of knowledge in the pre-

test. The percentages of words rated at the two initial levels of knowledge suggest that the control group reported slightly more entirely unknown words, which had a 0 score (66%), than did the experimental group (61%). Table 6-5 also shows that there was a small percentage difference in the number of words at the slight knowledge level (score 1). The experimental group reported having seen about 28% of the target words before, whereas the control group showed 22% knowledge at this level.

Interestingly, the data in the table above reveal that, overall, the two groups were at the same level for receptive and productive word mastery before the ER programme started. Each group showed a receptive knowledge of about 7% of the 1212 target words. At a relatively lower level, the two groups demonstrated productive knowledge of about 4% of the test items. These figures mean that both groups recognised only a small number of the target words, at these two levels of word knowledge.

Having explored the results from the two word categories individually, there appears to be no difference between the two groups of participants in terms of their initial off-list receptive and productive knowledge. Both groups equally knew about 6% of off-list words at a receptive level and 3% of these lexical items at a productive level of use. However, a small difference between the two groups did appear when the 2K items rated at receptive and productive levels were calculated. The control group showed relatively better receptive 2K words knowledge (11.5%) in comparison with the experimental group (8%). Conversely, the experimental group demonstrated slightly better productive knowledge (8.8%) of words from this category, than did the control group (5.5%).

This difference between the two groups at the receptive and productive levels appears not to be substantial as the overall results revealed very close figures. A Chi-square test, which measures whether there is a significant difference

between the effect of two categorical independent variables on a categorical dependent variable, was performed, to determine whether the experimental and control groups' pre-test overall word knowledge percentages (presented in Table 6-5) were distributed differently across the four levels of knowledge. The test failed to indicate a significant difference between the two groups, $X^2(3) = 1.09$, $P > 0.05$. This suggests that the slight difference for each of the four scale levels was not significant between the two groups, as when the observed and expected values were tested, it was indicated that the two groups had similar knowledge of the target words at the time of the pre-test.

6.3.2 Experimental vs. control groups' target word knowledge improvement

After the 18 experimental participants had finished the ER treatment, the same individualised pre-tests were again administered to them and the control group. To determine whether there was any overall improvement in the two groups' knowledge of the target words in the post-test, a Paired Samples t-test was performed, using the data from the participants' four test scores (0,1,2, & 3). Data for the two groups' results is presented separately below.

6.3.2.1 Overall incidental vocabulary gain

In order to compare the experimental and control groups' results in the first post-test, which was administered one week after the ER programme, I first checked whether my data satisfied the assumption of normality, in order to decide which approach to use.

Table 6-6: Tests of Normality

Word Category	Group	Kolmogrov-Smirnov test	
		Statistic	p-value
<i>Off-list</i>	Experimental	0.130	0.200
	Control	0.177	0.139
2K	Experimental	0.149	0.200
	Control	0.112	0.200
Total	Experimental	0.113	0.200
	Control	0.091	0.200

Table 6-6 above, displays the results of the tests to find normality of the data using the Kolmogrov-Smirnov test. At the 5% level of significance, the data sets are considered normal. Hence, a parametric Paired-Samples t-test was performed to compare the pre and post target word knowledge of the participants in both groups.

6.3.2.1.1 POST-TEST RESULTS

To assess the improvement in the learning of the target words by both groups, a Paired-Samples t-test was carried out to compare their scores in the first post-test.

Table 6-7: Comparison of the experimental and control group post-vocabulary test scores, assessed by means of the Paired Samples t-test

Group	N	Mean	SD	<i>t</i>	Sig (2-tailed)
<i>Experimental</i>	18	1.52	0.30	10.69	0.000
<i>Control</i>	18	0.49	0.11		

Mean score= 0-3

Table 6-7 reports the mean scores of the experimental and control groups in the post-test. When the two groups are compared, it seems clear that those who participated in the ER programme knew significantly more vocabulary than

those who only took the pre-and post-tests. The mean score for the experimental group (M= 1.52, SD= 0.30) was greater than the control group's post-test score (M= 0.49, SD= 0.11). As can be seen in the table above, for the experimental group, the difference was statistically significant; $t(18)=10.69$, $p<0.001$.

Further analysis was conducted by using a Paired-Sample t-test, to compare the difference in performance for each group of participants on the pre-and post-test.

Table 6-8: Mean comparison of the experimental and control groups on the pre- and post-tests.

Group	N	Pre-test		Post-test		<i>t</i>	Sig (2-tailed)
		Mean	SD	Mean	SD		
<i>Experimental</i>	18	0.52	0.14	1.52	0.30	-12.81	0.000
<i>Control</i>	18	0.47	0.13	0.49	0.11	-1.07	0.301

Mean score= 0-3

The results presented in Table 6-8 demonstrate the overall scores for the two categories of test items (Off-list and 2K). From this table, it emerges that, in comparison with the control group's mean score, which had slightly improved from the pre-test (M= 0.47, SD= 0.14) to the post-test (M= 0.49, SD= 0.11), the mean score for the experimental group in the pre-test (M= 0.52, SD= 0.14) had significantly increased to 1.52 (SD= 0.30) in the post-test. This data reveals that the difference between the mean score in the two tests was not significant for the control group ($p>0.05$). However, for the experimental group, the difference was statistically significant, as the p value was less than 0.05. These results show that, in comparison with the control group, the vocabulary learning of the experimental group in the two target categories of words had increased substantially after they had finished the ER treatment. Moreover, the performance of the two groups was much more variable, as

indicated by the very different standard deviations shown in Table 6-8 above. However, unlike the control group's decline in standard deviation (from 0.13 to 0.11) as a percentage of their means, the experimental group's standard deviation for the learning of target words increased after they had finished the reading programme (from 0.14 to 0.30), indicating that the experimental group had acquired a greater number of the target words.

To further explore the mean difference in of the level of knowledge of the target words shown in the in the pre-and post-test by the participants, a percentage of the words at each level in the two test administrations was analysed using a Paired-Samples t-test. Output data for each of the two group's participants are presented below.

Table 6-9: Mean percentage difference for words rated by the experimental group at each the four word knowledge scale levels in the pre-and post-test.

Knowledge Level	Test	N	Mean	SD	<i>t</i>	<i>Sig. (2-tailed)</i>
0* <i>(no knowledge)</i>	Pre-test	18	62.01	5.59	13.28	0.000
	Post-test	18	35.02	7.00		
1 <i>(slight knowledge)</i>	Pre-test	18	27.89	5.27	-0.94	0.361
	Post-test	18	30.84	13.02		
2 <i>(receptive knowledge)</i>	Pre-test	18	5.80	3.44	-8.59	0.000
	Post-test	18	17.51	7.41		
3 <i>(productive knowledge)</i>	Pre-test	18	4.30	4.15	-6.50	0.000
	Post-test	18	16.62	10.77		

*= score mark

Mean= percentages of the four word knowledge scores.

Table 6-9 displays the data for the experimental group. As can be seen, the table indicates a noticeable change in the experimental participants' mean percentages for three levels of knowledge of the target words (0, 2, &3) after they completed the ER programme. The experimental group participants' mean knowledge of items obtaining score 1 (*not really sure of the word meaning*) in the pre-test (M= 27.89, SD = 5.27) reflected a very small improvement in the

post-test ($M = 30.84$, $SD = 13.02$). The difference between the means at this level, however, was found not to be significant; $t(18) = 0.94$, $p > 0.05$.

On the other hand, their receptive knowledge (Score 2) mean shows that there was an improvement in the number of words for which the experimental group were able to provide Arabic equivalents in the post-test. The mean percentage at this level increased from 5.80 ($SD = 3.44$) in the pre-test, to 17.51 ($SD = 7.41$) in the post-test. The negative t value (8.59) indicates that the mean receptive knowledge of the experimental group after the reading programme was higher than their mean receptive knowledge before they started the programme. This significant difference is confirmed by the p value ($P < 0.001$).

Table 6-9 also shows that the productive knowledge (Score 3) of the experimental group improved noticeably, from 4.30 ($SD = 4.15$) in the pre-test, to 16.62 ($SD = 10.77$) in the post-test, $t(18) = 5.34$, $P < 0.001$. Therefore, it can be concluded that the mean scores for before and after the reading programme differed significantly ($P < 0.001$), in that the productive knowledge of the experimental group improved significantly over the nine weeks of the ER treatment, in comparison with that of the control group.

In looking at the words which were entirely unknown to the experimental group participants (Score 0) in the pre-test, it was not surprising to find a considerable decline in the mean of unknown words in the post-test ($M = 35.02$, $SD = 7.00$) in comparison with their mean in the pre-test ($M = 62.01$, $SD = 5.59$), $t(18) = 13.28$, $P < 0.001$. This significant difference was expected, since many words at this level in the pre-test had improved to a higher level of knowledge after the participants finished their ER treatment.

Up to this point, an analysis of the experimental group's pre and post-test knowledge rates has been presented. The control group's results are shown in Table 6-10 below.

Table 6-10: Mean percentage difference of words rated by the control group at each of the four word knowledge scale levels in the pre-and post-test.

Knowledge Level	Test	N	Mean	SD	<i>t</i>	<i>Sig. (2-tailed)</i>
0* (no knowledge)	Pre-test	18	67.44	8.33	1.93	0.070
	Post-test	18	64.73	5.65		
1 (slight knowledge)	Pre-test	18	21.56	6.93	-2.37	0.030
	Post-test	18	25.08	4.08		
2 (receptive knowledge)	Pre-test	18	7.24	3.63	1.72	0.104
	Post-test	18	6.70	3.08		
3 (productive knowledge)	Pre-test	18	3.76	3.28	0.88	0.394
	Post-test	18	3.48	2.77		

*= score mark

Mean= percentages of the four word knowledge scores

The output from a Paired-Samples *t*. test, presented in Table 6-10, compares the overall mean percentage of words rated by the control group at each level in the pre-and post-test. In comparison with the experimental group, who showed a significant improvement in their mean percentages in the post-test, particularly at the receptive and the productive levels, the difference between the pre-and-post-test percentages for the control group appeared to be less significant at these levels.

The only significant difference was found with words which had scored 1 (*seen it before but not sure of its meaning*). The mean percentage of words rated at this level in the pre-test (M= 21.56, SD= 6.93) rose to 25.08 (SD= 4.08) in the post-test. This, statistically, was the only significant difference between the other three levels of word knowledge as the *t* value (2.37) and *p* value ($P < 0.05$) indicate. This may be as a result of their having seen the words in the pre-test. The difference between the mean percentage of words with a 0 score “*haven't seen the word before*”, slightly decreased from 67.44 (SD= 8.33) in the pre-test, to 64.73 (SD= 5.65) in the post-test. This difference, however, was not significant, as the *p* value was greater than 0.05 ($t(18) = 1.93$).

The data revealed concerning receptive and productive knowledge levels, revealed similar findings, in that the output showed similar results for the number of words the control group rated at these two levels in the pre- and post-test. The mean score for receptive knowledge was 7.24 (SD= 3.63) in the pre-test, which slightly decreased to 6.70 (SD= 3.08) in the post-test, but statistically these two means did not significantly differ in the two tests, $t(18) = 1.72, P > 0.05$. Similarly, the mean percentage for words rated at the *productive* level showed that there was no significant difference between their pre- (M= 3.76, SD= 3.28) and post-test (M= 3.48, SD= 2.77) scores, $t(18) = 0.88, P > 0.05$.

When looking at the overall vocabulary gain difference between the experimental and control group in the post-test, the results indicate a superior growth of absolute word gain by the experimental group. As can be seen from Table 6-9 above, the mean percentage of words rated at the receptive and productive levels of knowledge in the post-test were significantly higher than the pre-test means for the experimental group. Table 6-10, in contrast, shows that the control group's mean percentages at these two levels slightly decreased in the post-test. This indicates that the control group had lost a very small amount of their receptive and productive knowledge in the post-test.

To consider this issue in a different way, the test scores by the two groups in the pre-and post-tests were compared by a cross-tabulation analysis. This was a necessary step, in order to calculate the percentages of target word knowledge growth shown by each group. This information is presented in Table 6-11.

Table 6-11: The experimental and control group's target word knowledge improvement

The Experimental Group					The Control Group						
<i>T1</i>	<i>T2</i>	<i>N</i>	<i>%</i>	<i>Change</i>	<i>T1</i>	<i>T2</i>	<i>N</i>	<i>%</i>	<i>Change</i>		
0	→	0	348	46.9	Stable	0	→	0	670	83.3	Stable
0	→	1	209	28.2	No gain	0	→	1	130	16.2	No gain
0	→	2	116	15.6	Gain	0	→	2	3	0.4	Gain
0	→	3	69	9.3	Gain	0	→	3	1	0.1	Gain
1	→	0	70	20.9	Attrition	1	→	0	102	37.9	Attrition
1	→	1	128	38.2	Stable	1	→	1	160	59.5	Stable
1	→	2	77	23.0	Gain	1	→	2	6	2.2	Gain
1	→	3	60	17.9	Gain	1	→	3	1	0.4	Gain
2	→	0	1	1.3	Attrition	2	→	0	5	5.5	Attrition
2	→	1	8	10.1	Attrition	2	→	1	10	11.0	Attrition
2	→	2	31	39.2	Stable	2	→	2	71	78.0	Stable
2	→	3	39	49.4	Gain	2	→	3	5	5.5	Gain
3	→	0	1	1.8	Attrition	3	→	0	2	4.2	Attrition
3	→	1	0	0.0	Attrition	3	→	1	4	8.3	Attrition
3	→	2	3	5.4	Attrition	3	→	2	4	8.3	Attrition
3	→	3	52	92.9	Stable	3	→	3	38	79.2	Stable

T1=Pre-test, T2=Post-test

The cross-tabulation analysis output in Table 6-11 shows the difference in the number of words acquired by each group. In order to statistically determine any significant post-test knowledge improvement or attrition for words rated 0, 1, 2, and 3 in the pre-test, a Chi-square test was performed to compare the experimental and control groups. The test revealed that the two groups of participants in the post-test significantly differed in their recognition of the words they rated as totally unknown in the pre-test (i.e. 0→0, 0→1, 0→2, 0→3), $X^2 = (3) = 47.8$, $p < 0.001$. Similarly, in a comparison between the experimental and the control groups' change in knowledge of words they reported as "have seen before" in the pre-test to their knowledge in the post-test, a Chi-square test showed that, there was a significant difference between the two groups' change in word knowledge at this level in the post-test (1→0, 1→1, 1→2, 1→3), $X^2 = (3) = 121.5$, $p < 0.05$.

When a Chi-square test was performed, a further statistical difference between the two groups appeared in the number of words the participants in the two

groups had rated at the receptive level in the pre-test, compared with their rating of these words in the post test. There was significantly different post-test knowledge between the two groups on these words (2→0, 2→1, 2→2, 2→3), $X^2 = (3) = 44.2$, $P < 0.001$. Unsurprisingly, a Chi-square output revealed that there was no statistically significant difference between the two groups on words they knew at the productive level in the pre-test (3→0, 3→1, 3→2, 3→3), $X^2 = (3) = 6.1$, as the p value was greater than 0.05. This was perhaps due to the participants' high level of mastery of these words, which enabled the two groups to retain most of these words at the same level of knowledge.

The Chi-square test results can be clearly seen in the experimental and control groups' word knowledge percentages (gain, attrition, stable) in Table 6-11 and, from this, it is noticeable that exposure to the target words through the ER materials had a strong effect on the experimental group's performance in the post-test. The data shows both the receptive and productive knowledge (score 2 & 3) of the experimental group improved after nine weeks of the extensive reading treatment. Overall, the experimental group demonstrated a receptive and productive gain of 31.2% of the unknown target words (361/1156: 56 words were found to be productively known by this group in the pre-test). This is demonstrated by 0→2, 0→3, 1→2, 1→3, 2→3. In contrast, the data indicates that the control group learnt only a small portion of 1.4% of the unknown target words (15/1164: 48 words were recognised at the productive level by the group in the pre-test). Interestingly, before being improved to the productive mastery stage in the post-test, a third of these words (5/15) had already been receptively known by the control group in the pre-test.

Table 6-11 reveals another interesting finding. The cases of attrition suggest that vocabulary knowledge is not stable and even productive word knowledge may decrease over time. However, a comparison between the ER group and the control group shows that free reading can be an effective means of maintaining knowledge of already known words (at receptive and productive levels). This

can be clearly seen in the cases of attrition (i.e. 2→0, 2→1, 3→0, 3→1, 3→2) within the control group results (25/1212), which were double those reported by the experimental group (13/1212).

Moreover, with the exception of the target words which were rated at the productive level (as further improvement was not possible), data from the control group's results demonstrate that knowledge level of over three quarters of the words (77.4%; 901/1164) remained stable (i.e. 0→0, 1→1, 2→2) in the pre-and-post-test. The experimental group, however, had less than half of these words (43.9%; 507/1156) remaining stable in the two tests.

Therefore it seems clear that the word knowledge of the experimental group tended to improve to a higher level. Not only did the participants' receptive and productive knowledge improve after the programme, some of the words they reported as entirely unknown in the pre-test (score 0) also developed to partial knowledge (i.e. 0→1) in the post-test (209/742: 28.2%). In contrast, the control group in the post-test reported partial enhancement of only 16.2 % (130/804) of the words they had rated as new in the pre-test. Consequently, it appears that the experimental group had better knowledge enhancement even at the very basic stage of word recognition, adding more support to the claim that the target vocabulary size improvement of these participants resulted from the reading treatment. Moreover, the above table shows that the experimental group improved their receptive and productive mastery of words they had reported as seen before in the pre-test (1→2= 23% and 1→3=17%) in comparison with words they did not know at all (0→2=15% and 0→3=9.3%). This gives an indication of the importance of ER in enriching knowledge of partially known words.

Up to this point, the overall gains by the experimental and the control group have been presented. However, further explanatory steps of analysis are required, to ascertain whether the selection of the target items from two

different frequency categories affected the test performance of the participants in both groups. The results for the off-list words are presented first.

6.3.2.2 Off-list words gain

Table 6-12: Comparison of the experimental and control group for post-test off-list item scores, assessed by means of the Paired-Samples t-test

Group	N	Mean	SD	<i>t</i>	Sig (2-tailed)
<i>Experimental</i>	18	1.12	0.29	12.92	0.000
<i>Control</i>	18	0.37	0.13		

Mean score (0-3)

Table 6-12 displays the results from a Paired-Samples t-test, which was performed to determine the difference between the experimental and control groups' off-list item mean scores. The test revealed that the mean for the experimental group ($M= 1.12$, $SD= 0.29$) was significantly greater than the mean for the control group ($M= 0.37$, $SD= 0.13$) in the post-test, $t=12.92$, $P<0.001$. Since the p-value of the off-list items in the pre-test (see Table 6-3) indicated no significant difference between the two groups, it can be concluded, as the mean score in Table 6-12 shows, that the experimental group's off-list words knowledge grew significantly after they had been exposed to the ER materials.

Table 6-13: Comparison of the experimental and control group levels of knowledge of off-list items in the pre-and post-test

Knowledge Level	Group	Pre-test		Post-test		<i>t</i>	Sig (2-tailed)
		M	SD	M	SD		
0* (no knowledge)	Experimental	67.98	6.29	36.22	7.33	12.17	0.000
	Control	74.15	9.59	73.13	7.01	0.67	0.457
1 (slight knowledge)	Experimental	24.39	6.75	30.64	14.38	-1.73	0.102
	Control	17.73	8.10	19.83	5.67	-1.38	0.186
2 (receptive knowledge)	Experimental	5.28	3.53	17.57	8.49	-8.11	0.000
	Control	5.21	3.36	4.13	2.96	2.30	0.035
3 (productive knowledge)	Experimental	2.35	2.65	15.57	10.69	-5.93	0.000
	Control	2.91	3.30	2.91	3.14	0.01	0.991

*=score mark

Mean= percentages of the four word knowledge scores

Table 6-13 contains the output from a paired-samples t-test, which was performed to determine the difference in the mean percentage of off-list words rated by the two groups at each of the four levels of knowledge in the pre-and post-test. Again, the data prove the clear superiority of the experimental group over their corresponding partners in the control group. There was a remarkable improvement of the experimental group's mean off-list word knowledge percentage at both the receptive and productive levels. As shown in Table 6-13 above, the increase of their mean knowledge of off-list words at the receptive level (Score 2) was highly significant, from 5.28 (SD= 3.53) in the pre-test, to 17.57 (SD= 8.49) after the intervention, $t(18)= 8.11$, $P<0.001$. In contrast, the control group mean percentage at this level declined from 5.21(SD= 3.36) in the pre-test to 4.13 (SD= 2.96) in the post-test. This drop in off-list knowledge was found to be statistically significant as the p value was less than 0.05, $t(18)=2.30$.

When examining mean percentages of both groups at the productive level (Score 3), similar findings emerged from a Paired-Samples t-test. The

experimental group's mean for off-list productive knowledge increased from 2.35 (SD= 2.65) in the pre-test to 15.57(SD= 10.69) in the post-test, which, as the statistics indicate, is statistically significant, $t(18)= 5.04$, $P<0.001$. For the control group, however, mean percentages at the productive level of knowledge remained steady in the pre (M= 2.91, SD= 3.30), and post-test (M= 2.91, SD= 3.14) indicating that there was no significant difference between the control group's off-list productive knowledge in the two test settings, $t(18)=0.01$, $P>0.5$.

There was a decline in the mean percentage of words rated as "*I haven't seen the word before*" in the post-test. This decline was statistically greater in the experimental group's mean, which decreased from 67.98 (SD= 6.29) in the pre-test to 36.22 (SD= 7.33) in the post-test, $t(18)= 12.17$, $P<0.001$. However, for the control group, the decline in the mean percentage of unknown words from the pre-test (M= 74.15, SD= 9.59) to the post-test (M= 73.13, SD= 7.01) was not statistically significant, as the p value was greater than 0.05, $t(18)=0.67$. This indicates that many of the entirely unknown off-list words were recognised at a higher level of word knowledge in the post test, by the experimental group, in comparison with the control group, whose knowledge remained the same.

Table 6-13 also indicates that the improvement of the two groups' off-list knowledge at the "*I have seen this word before, but I am not sure of its meaning*" level (Score 1) did not significantly differ in the pre-and post-test. The mean percentage for the experimental group was 24.38(SD= 6.75) in the pre-test, which rose to 30.64 (SD= 14.38) in the post-test, whereas the mean for the control group slightly increased from 17.73 (SD=8.10) to 19.83 (SD=5.67). Statistically, the two groups' knowledge at this level was not significant, as the p value in the two groups was greater than 0.05 (for the experimental group $t(18)=1.73$, $p=0.102$; for the control group $t(18)=1.38$, $p=0.186$).

In order to elaborate on the statistically significant difference between the two groups on off-list gain, which is presented in Table 6-12 above, a cross-tabulation analysis was used to measure their knowledge improvement for this category of words. This is shown in Table 6-14 below.

Table 6-14: The experimental and control group's knowledge improvement for off-list words

The Experimental Group					The Control Group						
<i>T1</i>	<i>T2</i>	<i>N</i>	<i>%</i>	<i>Change</i>	<i>T1</i>	<i>T2</i>	<i>N</i>	<i>%</i>	<i>Change</i>		
0	→	0	251	45.5	Stable	0	→	0	524	88.2	Stable
0	→	1	159	28.8	No gain	0	→	1	68	11.4	No gain
0	→	2	86	15.6	Gain	0	→	2	1	0.2	Gain
0	→	3	56	10.1	Gain	0	→	3	1	0.2	Gain
1	→	0	44	22.6	Attrition	1	→	0	62	41.3	Attrition
1	→	1	63	32.3	Stable	1	→	1	85	56.7	Stable
1	→	2	47	24.1	Gain	1	→	2	2	1.3	Gain
1	→	3	41	21.0	Gain	1	→	3	1	0.7	Gain
2	→	0	0	0.0	Attrition	2	→	0	2	4.4	Attrition
2	→	1	3	6.4	Attrition	2	→	1	8	17.8	Attrition
2	→	2	22	46.8	Stable	2	→	2	33	73.3	Stable
2	→	3	22	46.8	Gain	2	→	3	2	4.4	Gain
3	→	0	0	0.0	Attrition	3	→	0	1	3.8	Attrition
3	→	1	0	0.0	Attrition	3	→	1	2	7.7	Attrition
3	→	2	0	0.0	Attrition	3	→	2	1	3.8	Attrition
3	→	3	21	100	Stable	3	→	3	22	84.6	Stable

% within the pre-test

Again, a Chi-square test was used on the off-list words for which participants obtained 0, 1, 2, and 3 in the pre-test, to compare the significant change in word knowledge in the post-test. The off-list test results in Table 6-14 reveal similar overall gain results to those in Table 6-11 presented earlier. The significant difference between the two groups' pre-and-post-test was determined by comparing the number of words which initially obtained 0 score in the pre-test (i.e. 0→0, 0→1, 0→2, 0→3), $X^2 = (3) = 267.6$, $p < 0.001$; words that gained 1 score in the pre-test (i.e. 1→0, 1→1, 1→2, 1→3), $X^2 = (3) = 81.3$, $p = < 0.001$; and those that were receptively understood by the two groups in the pre-test (i.e. 2→0, 2→1, 2→2, 2→3), $X^2 = (3) = 23.1$, $p < 0.001$. A Chi-square test, however, failed to indicate a significant difference between the

two groups' post-test results for the words that both knew at the productive level of use in the pre-test (i.e., 3→0, 3→1, 3→2, 3→3), $X^2 = (3) = 3.5$, which means that the p value was greater than 0.05).

The data for the control group in the table above show very small (0.9 %) receptive and productive gains (0→2, 0→3, 1→2, 1→3, 2→3) for the off-list words, in comparison with the experimental group's knowledge enhancement in the post-test (252 of 742 off-list with pre-test score 3 exclusion). The vast majority of words in this category (81.5%) remained unchanged (0→0, 1→1, 2→2, 3→3) in the control group's pre-and post-test results. In contrast, the experimental group's results revealed a much greater percentage (32%) of off-list receptive and productive gains. The data also show that, compared to the experimental group's level of attrition (17.8%), the number of cases for the control group were considerably higher (up to twice the amount, 32.1%).

6.3.2.3 2K words gain

The same analysis methods were applied to words from the 2K category. A Paired-Samples t-test was again conducted, to compare the difference between the two groups' mean scores for 2K items in the post-test.

Table 6-15: Comparison of the experimental and control group on post-test 2K items scores assessed by means of the Paired-Samples t-Test

Group	N	Mean	SD	<i>t</i>	Sig (2-tailed)
<i>Experimental</i>	18	1.22	0.38	5.61	0.000
<i>Control</i>	18	0.74	0.19		

Mean score (0-3)

Table 6-15 revealed that in the post test, the mean for the experimental group (M= 1.22, SD 0.38) was significantly greater than the mean for the control group (M=0.74, SD=0.19). Although a pre-test comparison (see Table 6-4)

revealed no significant difference between the two groups' 2K words knowledge, this difference was statistically significant in the post-test as the p value was less than 0.001, $t=5.61$. The data suggest that the experimental group's 2K words knowledge had increased significantly after they had finished the reading programme.

This difference can be further explored by a comparison of the mean percentages of 2K items rated by the two groups of participants at each of the four levels of knowledge in the pre-and post-test.

Table 6-16: Comparison of the experimental and control group levels of knowledge 2K items in the pre-and post-tests

Knowledge Level	Group	<i>Pre-test</i>		<i>Post-test</i>		<i>t</i>	Sig (2-tailed)
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
0* (no knowledge)	Experimental	49.77	14.86	32.58	11.51	4.87	0.000
	Control	53.70	14.31	47.52	11.07	2.43	0.027
1 (slight knowledge)	Experimental	35.07	12.86	31.28	15.38	0.98	0.341
	Control	29.39	13.58	35.87	8.31	-2.29	0.035
2 (receptive knowledge)	Experimental	6.88	5.97	17.34	9.75	-4.25	0.001
	Control	11.44	6.01	11.98	6.22	-0.61	0.550
3 (productive knowledge)	Experimental	8.27	8.92	18.80	13.94	-6.28	0.000
	Control	5.47	6.42	4.63	4.79	1.29	0.215

*=score mark

Mean= percentages of the four word knowledge scores

The data reported in Table 6-16 above show a significant difference between the two groups' 2K word knowledge in the post-test. The experimental group appears to have gained substantial growth in their 2K word knowledge, when compared with the control group. This was particularly evident at the receptive and productive levels. As can be seen from the table, the experimental group's mean percentage for receptive knowledge grew significantly from 6.88 (SD=5.97) in the pre-test to 17.34 (SD=9.75) in the post-test ($t(18)=4.25$,

$P < 0.01$), whereas the mean receptive knowledge for the control group remained mostly the same (pre-test, $M = 11.44$, $SD = 6.01$; post-test $M = 11.98$, $SD = 6.22$), as the p value shows no significant difference ($t(18) = 0.61$, $P > 0.05$).

A similar result was gained when looking at the productive level of 2K knowledge. The experimental group showed a greater improvement in their mean percentage of knowledge, from 8.27 ($SD = 8.92$) to 18.80 ($SD = 13.94$), $t = 6.28$, $P < 0.001$, in comparison with the control group's mean which slightly decreased from 5.47 ($SD = 6.42$) in the pre-test to 4.63 ($SD = 4.79$) in the post-test, $t(18) = 1.29$, $P > 0.05$.

The data from items rated at the "*I haven't seen the word before*" level showed a clear decline of the mean percentage in the post-test by both groups. While the experimental group mean for entirely unknown words in the pre-test ($M = 49.77$, $SD = 14.86$) dropped to 32.58 ($SD = 11.51$) in the post-test, the control group mean percentage in the pre-test ($M = 53.70$, $SD = 14.31$) fell to 47.52 ($SD = 11.07$). The decrease was statistically significant, as the p value in both groups was less than 0.05. The clear reduction in the number unknown 2K words implies improvement of word knowledge in the post-test.

Interestingly, the control group showed better growth in their mean percentage of 2K words knowledge at the "*I have seen the word before but not sure of its meaning*" level of knowledge, which was 29.39 ($SD = 13.58$), but improved to 35.87 ($SD = 8.31$) in the post-test. The Paired-Samples t -test indicated that the difference between the two means was statistically significant, $t(18) = 2.29$, $P < 0.05$. Conversely, at this level of knowledge, the experimental group's mean decreased from the pre-test ($M = 35.07$, $SD = 12.86$) to the post-test ($M = 31.28$, $SD = 15.38$). This difference between the two means, however, was not statistically significant, as the p value for the paired t -test was greater than 0.05 ($t(18) = 0.98$).

To explore these statistical findings further, a cross-tabulation analysis was also conducted on the 2K words, to measure the amount of newly acquired knowledge gained by the experimental and the control groups.

Table 6-17: The experimental and control group's gain of 2K words

The Experimental Group					The Control Group						
<i>T1</i>	<i>T2</i>	<i>N</i>	<i>%</i>	<i>Change</i>	<i>T1</i>	<i>T2</i>	<i>N</i>	<i>%</i>	<i>Change</i>		
0	→	0	97	51.1	Stable	0	→	0	146	69.5	Stable
0	→	1	50	26.3	No gain	0	→	1	62	29.5	No gain
0	→	2	30	15.8	Gain	0	→	2	2	1.0	Gain
0	→	3	13	6.8	Gain	0	→	3	0	0.0	Gain
1	→	0	26	18.6	Attrition	1	→	0	40	33.6	Attrition
1	→	1	65	46.4	Stable	1	→	1	75	63.0	Stable
1	→	2	30	21.4	Gain	1	→	2	4	3.4	Gain
1	→	3	19	13.6	Gain	1	→	3	0	0.0	Gain
2	→	0	1	3.1	Attrition	2	→	0	3	6.5	Attrition
2	→	1	5	15.6	Attrition	2	→	1	2	4.3	Attrition
2	→	2	9	28.1	Stable	2	→	2	38	82.6	Stable
2	→	3	17	53.1	Gain	2	→	3	3	6.5	Gain
3	→	0	1	2.9	Attrition	3	→	0	1	4.5	Attrition
3	→	1	0	0.0	Attrition	3	→	1	2	9.1	Attrition
3	→	2	3	8.6	Attrition	3	→	2	3	13.6	Attrition
3	→	3	31	88.6	Stable	3	→	3	16	72.7	Stable

% within the pre-test

Table 6-17 reveals similar results to those for the off-list words. From the percentages of word gains, it is clear that the experimental group achieved higher results than the control group in this category too. In order to measure the significance of this difference, a Chi-square test was used on the words which were marked 0, 1, 2 and 3 in the pre-test. Again, the test showed no significant difference between the two groups for words already known at the productive level in the pre-test (3→0, 3→1, 3→2, 3→3), $X^2 = (3)=4.0$, which indicated that the p value was greater than 0.05. The difference between the two groups, in the number of 2K words with scores of 0, 1, and 2 in the post-test was again found to be statistically significant, when a Chi-square test was run: (0→0, 0→1, 0→2, 0→3), $X^2 = (3)=47.8$, $P<0.001$; (1→0, 1→1, 1→2,

1→3), $X^2 = (3)=41.1$, $P<0.001$, (2→0, 2→1, 2→2, 2→3), $X^2 = (3)=28.4$, $P<0.001$.

In addition, the ER participants showed both a receptive and productive gain (0→2, 0→3, 1→2, 1→3, 2→3) of 30.2% of unknown 2K words (109/ 362). However, the control group demonstrated a gain of only 2.4% (9/397) in this category. Again, the control group's word knowledge attrition (13%) was slightly higher than that of the experimental group (9%). Moreover, the amount of pre-test vocabulary with scores of 0, 1, and 2 and which gained the same marks in the post-test (i.e. 0→0, 1→1, 2→2) was greater for the control group (69.2%) than for the experimental group (50.9%).

6.3.3 Summary of main findings

A comparison of the two groups' performance in the vocabulary pre-test revealed several interesting points. The findings in Table 6-2 show that, overall, for the two groups, the mean score for pre-existing target word knowledge was not significantly different ($p>0.05$). This suggests that participants did not differ in their knowledge of the words from the two frequency levels of use in English (2K and off list). It is interesting to note that, as shown in Table 6-5, the vast majority of the target words were rated as partially, or fully unknown by the participants. The participants in both groups seemed to know a very small number of words at the receptive and productive levels of knowledge. Nevertheless, it was important to omit words from the results analysis which were already known at the productive level, as they had no further chance of improvement through extensive reading. Improvements in knowledge for both groups were possible for about 95% of the words at receptive or productive levels after reading. Participants in both groups rated the vast majority of words as: no knowledge, slight knowledge and receptive knowledge. This similarity in the knowledge of both groups suggests that the procedures for selecting the target words, as well as those which were applied during the paired-sampling techniques, as were discussed earlier in the

Methodology chapter, proved to be very helpful in this study. Since the matching of the experimental and the control participants was based on the closeness of their marks on the VLT, it is likely that their knowledge of the target words was also close also. Other possible interpretations for such close similarities between the two groups' prior knowledge of target words from the two different frequency levels are likely to be due to the participants' similar English language learning experiences and the amount of language instruction they had received.

A study of the relevant key findings in the pre-test exposed a further interesting point. To a certain extent, it was expected that the vast majority of words from the off-list category would not be recognised by the two groups because of their low frequency of use in English language. This was proved to be true, as the results in Table 6-5 indicate, since the experimental and control groups both knew about 3% of the items at a productive level and recognised 5.8% and 5.5% respectively, at the receptive level. The striking result, however, is that a high proportion of the lexical items selected from the higher frequency level of use (2K) were also found not to be familiar to the participants in the two groups. The experimental group rated about 48% and the control group 53% of these words at a level of entirely *unknown* (0 score). Moreover, the groups rated about 35% and 30% (respectively) of the items in this frequency band, as *seen before, but I am not sure of their meaning* (score 1). Overall, each group recognised only a small percentage of these words at either the receptive or productive levels of use. These findings suggest that the piloting procedures for selecting target words from words within this frequency were very effective. As stated earlier, in the methodology chapter, although these words belong to the second thousands of most frequent words used in English, the pilot study revealed that 2K words used in the present assessment had been found unfamiliar to the pilot subjects. Therefore, it should not be surprising that participant in the main study did not recognise many of these words.

One aim of this investigation was to explore the extent to which an ER programme can enhance the vocabulary knowledge of Libyan, adult, EFL learners. The general hypothesis tested, was that the ER participants, who were exposed to a variety of interesting graded readers, would show greater incidental vocabulary gains than was usual for such learners. Despite the fact that reading contributes to multiple aspects of vocabulary knowledge, very little attention has been given to the measurement of knowledge aspects other than meaning. Therefore, word knowledge in this study was measured using a four-level knowledge scale instrument, which ranged from no knowledge to productive knowledge of a word. In order to assess gains at each of the levels of word use, absolute learning at the receptive and productive levels were specifically examined.

Findings showed that ER really could improve the Libyan, EFL learners' vocabulary size. Although the ER treatment lasted for only nine weeks, the results were clearly significant and showed that the participants were able to learn about a third of the target words at both receptive and productive levels ($p < 0.05$). It was found that the treatment group had, on average, picked up 361 words, after a nine week ER programme. This suggests that one in every 3.2 tested words was learned incidentally from the experimental graded books. This included an improvement of 39 words (11%) from the receptive to the productive level of mastery. In contrast, the control group, who did not participate in the reading treatment, learned a very small number of the two frequency-band target words (1.4%). As a result of the careful procedures implemented in selecting the target words, such insignificant improvement by the control group ($P > 0.05$) clearly suggests that the learning by the experimental group resulted from incidental acquisition via the ER programme.

If it is assumed that the treatment participants read books on all six days of each experimental week, the gains of 361/1164 words indicate that they learned, on average about 6.68 words, at receptive and productive, levels each

day. These numbers seem encouraging if, as discussed in the literature review chapter, it is taken into consideration that, in general, an L1 student annually learns an approximate number of 1000 words, out of an average number of one million words read (Goulden et al., 1990; Nagy, 1997). The gain rate revealed by the current investigation suggests that the Libyan, EFL students would be able to meet this target number, if they were involved in ER during their university semesters.

Many of the previous studies (e.g., Saragi et al., 1978; Pitts et al., 1989; Day et al., 1991; Dupuy & Krashen, 1993) found far smaller incidental gains than have been reported by the present study. Horst et al., (1998) in reviewing these studies, estimated an incidental pick up of roughly one word in each dozen tested. Nevertheless, the comparison between incidental vocabulary gain rates in the current study and the aforementioned studies seems meaningless. One important factor that was present in this study, that has been absent in most of the previous L2 incidental vocabulary learning research, was the specification of the reading task. The conditions needed for ER are considerably different from those used in most of the previously mentioned studies. For extensive reading, the subjects should read at their own language level. In contrast, in most of those studies, the researcher selected the reading text and gave it to their subjects to read, without control over its level of difficulty or interest to the subjects. This, of course, poses a question about the learners' reading comprehension as, in one of these studies, half of the subjects found it difficult to read two chapters of *A Clockwork Orange* (Pitts et al., 1989).

Therefore, a comparison of the results from the current study, to the incidental vocabulary rates reported in past studies using the ER approach, is warranted. Although Al-Homoud and Schmitt (2009) implemented an entirely different testing approach for measuring word knowledge improvement, their findings (5.75 words learning a day during 10 weeks of extensive reading) are closely congruent with the rate noted in the present study (6.68 per day). The given

subjects in the two studies read various numbers of books, at their level of language proficiency, that also met their interest, and it is likely therefore, that they encountered the target words in a passably comprehensible input. If Krashen's (2004) Comprehension Hypothesis is to be considered here, the greater amount of incidental gain in these studies should not be surprising, as the results seem in agreement with Krashen's (2004) claim that acquisition of language usually takes place when learners understand what they read.

As mentioned earlier, the present study replicated most of the methodological procedures that Horst (2005) implemented in her pilot research (i.e., the book scanning and building up of a corpus, the target word sampling procedures, the reading approach, the length of the treatment and the individualisation of the testing). What is striking, however, is that the findings in the two studies reveal considerably different results, in terms of the amount of incidental off-list word gains that were reported. The 18 partial and full increases which Horst's subjects acquired from 35 (51%) off-list words after six weeks of ER (about one in every two words tested), was higher than the one to three receptive and productive improvements found in the current study (32% = 252/742 off-list). If it was assumed that the off-list words in both studies had no chance of being encountered elsewhere, the cause of the difference is clearly related to the slight change in the research design for the two studies. One reason that might have allowed Horst's subjects to learn a larger proportion of the off-list words, is the way the ER was implemented. The subjects of the present study undertook their reading solely in the experimental library without permission to use a dictionary. Instead, they were encouraged to use guess work to obtain the overall meaning of the texts, although some participants reported glossary use. In contrast, in Horst's study, the subjects were provided with weekly one-hour instruction on reading, and they were allowed to read the books at home. They may have performed some vocabulary exercises and used dictionaries to practise and to find out the meanings of unfamiliar words. The assumption that learners may report higher rates of learning, if dictionaries are used, is supported by Cho and Krashen's (1994) research. They found that the subjects

who used a dictionary consistently whilst reading, learned as many as 50% of the unknown words (17/34), whereas those who did not use a dictionary at all, learned just 7 of the 34 target words (21%). For these reasons, the present study seems more precise, in reflecting typical incidental word learning, as it coincides more closely with normal ER conditions.

There are other reasons which may also lead to an argument that the findings from this study are more reliable than Horst's (2005) results. For instance, she tested words that appeared in some parts, but not all parts, of the texts. The scanning of all the words in the target books within the current study allowed for more precise measurement of the incidental increase in word gain that each participant encountered during the experiment. Moreover, there was a successful administration of the same individualised pre-and post-test items and format, but this was not the case in Horst's study, because her findings were based on results from two different pre and post-test formats. That is, in the post test she used a four-level knowledge scale to test the learning of words which her subjects had ticked "NO" on the checklist test used for testing their pre-existing knowledge of the target words. Waring and Takaki (2003) found that vocabulary gain results can be greatly affected by the test format used in the research, which suggests that Horst's results could have been different if she had given her subjects the same test in the two administrations. Furthermore, in Horst's study, learning for each participant was assessed by testing of a smaller number of off-list words. The inclusion of a far larger number of target words in the individualised vocabulary tests could have served to provide an opportunity for the participants to show a greater learning rate.

Other interesting findings emerge when levels of acquired knowledge are considered. Of the 361 words, which the experimental group in this study learned from the extensive reading, they were able to use nearly half in a productive manner (47%: 169 words). A number of researchers (Laufer, 1998;

Laufer & Paribakht, 1998; Fan, 2000) have estimated that the majority of language learners acquire a larger receptive than productive vocabulary. However, Melka (1997) surveyed a number of studies and concluded that there was a very small gap between the learners' receptive and productive knowledge size and, overall, learners knew 92% of their receptive vocabulary at the productive level. Findings in the current study seem to be in agreement with the estimations made by Fan (2000), Laufer (1998) and Laufer and Paribakht, (1998), but differ greatly from Melka's (1997) study. The difference seems to lie in what aspects of word receptive and productive knowledge are being investigated, and the degree of complexity of the measuring instruments that are used in testing the learners' word knowledge, at different levels. There is always a chance of guessing on multiple-choice receptive tests, whilst a productive test is more demanding, requiring participants to demonstrate more aspects of knowledge than a receptive format. This aspect of the test is very important. Since productive word learning in the present study was measured by the learners' ability to use the word appropriately in a sentence, it is likely that the participants found it more challenging to use the word productively than to match the words to definitions or to provide an L1 translation, in order to demonstrate their receptive knowledge of use. Webb (2008b), for instance, found that subjects tended to obtain higher scores and demonstrate greater knowledge on tests which measured receptive word learning, than on the productive test, which asked for cued recall.

Additional insights can be obtained by studying the results in terms of their level of knowledge (Table 6-11). The results indicate that the treatment group obtained larger proportions of receptive and productive development of the words they had previously reported at Level B (score 1) "I have seen the word before, but I am not sure of its meaning," in comparison with their knowledge improvement at the same two levels of mastery of words they had previously rated at Level A (score 0) "*I have not seen the word before.*" If the claim about the incremental nature of vocabulary acquisition (Nagy et al., 1985) is accepted

and any movement towards mastery is important, then these findings suggest that ER can be effective in promoting this process.

One point to note in the case of this study is that, because of the individualised nature of extensive reading, it was not feasible to administer the post-test immediately after the first participant had finished reading, as the participants finished reading their selected books at varied times within the final week of the experiment. However, most of the studies reviewed earlier in the literature review reported incidental gain from conditions in which all of the subjects had finished at a similar time before they took the post-test. Of course, the difference in the time that each of the participants finished her/his reading may have affected the results of the present study. For instance, some of the participants finished their reading roughly 10 days before the post-test, while others finished few days later. Indeed, evidence from previous research showed a decline in mean knowledge scores just one week after the subjects took their immediate post-test (Waring and Takaki, 2003). Memory research also indicates that most word knowledge attrition takes place shortly after the learning event and then declines over time (Baddeley, 1990). This suggests that some participants in the current study may have picked up more vocabulary than was reported in the above studies. Due to the fact that they finished their reading relatively earlier than other participants in the reading programme, it is possible that some of the new knowledge that they had gained from the stories could have been forgotten by the time they took the post-test.

A comparison between the two groups' performances in the pre-and post-tests suggests that the experimental group picked up their new knowledge of the target words exclusively from the ER graded books. This is because, the results of their corresponding control partners, who did not take part in the reading programme, showed that they had only gained a less than 2% increase in their receptive and productive knowledge of the target words, in comparison with the pre-test. A statistical comparison between the mean gain of the two groups

in Table 6-7 yielded a difference at a very high degree of significance ($p > 0.05$), showing that the experimental group's incidental knowledge gain was superior to that of the control group. Clearly, this difference lies in the effects of the ER programme. However, the potential existence of other factors, which could have led to such small gains in the number of known 'new' target words by the control group, should not be denied. As an example, the reduction in normal classroom instruction, due to the situation in Libya¹³ at the time when this experiment was conducted, seems to have had an impact on the students' EFL learning. It appears that the control group had very limited exposure to the English language outside of what has traditionally been the sole source of input (the university classroom), while their experimental corresponding partners were exposed to the English graded reader books. Even with the inclusion of items selected from the high frequency word category (2K), which were supposed to have increased the opportunities for encounters outside the reading treatment, the control group showed a very small and insignificant improvement in their knowledge in the post-test (see Table 6-16).

6.4 Research Question 2: Does word repetition affect receptive and productive incidental vocabulary learning from extensive reading? If so, how many exposures are needed for incidental vocabulary learning?

In order to answer this research question, the target words were organised into eight sub-groups according to the number of times the target word were repeated in the stories. The sub-grouping technique was necessary here for two reasons. Firstly, the study was investigating a very large number of words,

¹³ Libya underwent a major change when long-term leader Muammar Gaddafi was toppled in a popular uprising in February 2011. The uprising started while this research was being conducted. It led to a ten-month armed conflict between forces loyal to Colonel Ghaddafi and those seeking to oust his regime. Like other government sectors, Educational institutions were affected by the situation there.

which occurred at very different levels of frequency. Therefore, it was not possible to study these frequencies exclusively. Second, there were many occasions where a very small number of words had a particular number of repetitions, so sub-grouping was an alternative and practical method for obtaining a consistent number of words to allow for a meaningful analysis of each of the eight frequency categories. As mentioned in the methodology chapter, words with a single encounter were excluded from this research. Therefore, the 1212 target words were arranged into 8 subgroups: 2-5 occurrences, 6-9 occurrences, 10-14 occurrences, 15-19 occurrences, 20-29 occurrences, 30-39 occurrences, 40-49 occurrences, and over 50 occurrences .

6.4.1 Receptive knowledge improvements

Table 6-18 : Receptive gain percentages according to each frequency groups

Number of occurrences	Improvement	Number of words	%
2-5 (159)	0 → 2	7	5.7
	1 → 2	2	
	Total	9	
6-9 (314)	0 → 2	13	8.0
	1 → 2	12	
	Total	25	
10-14 (230)	0 → 2	31	24.3
	1 → 2	25	
	Total	56	
15-19 (98)	0 → 2	19	32.7
	1 → 2	13	
	Total	32	
20-29 (124)	0 → 2	29	30.6
	1 → 2	9	
	Total	38	
30-39 (56)	0 → 2	8	26.8
	1 → 2	7	
	Total	15	
40-49 (50)	0 → 2	7	22.0
	1 → 2	4	
	Total	11	
50+ (46)	0 → 2	2	15.5
	1 → 2	5	
	Total	7	
All words (1077)	0 → 2	116	17.9
	1 → 2	77	
	Total	193	

Table 6-18 displays the percentage of words, in each of the eight frequency categories, which were receptively acquired. In order to calculate only the actual receptive gains, words which the participants had recognised at receptive and productive levels in the pre-test, were excluded from this calculation. This was because there was no chance of their knowledge of these words improving to the receptive level in the post-test. As can be seen from the table, the figures show that words in the first two frequency categories (2-5, 6-9) had less than 10% probability of receptive gain. Noticeably, nevertheless, there was a gradual rise in the percentage of words gained, from 5.7% in the 2-5 repetition group, reaching a peak at 32.7% after 15-19 exposures. One point of interest in the table above is that while participants knew the meanings of just 8% of words after 9 occurrences, the participants' receptive knowledge gain jumped remarkably, to three times as many words (24%) after 10 encounters, equalling just under a quarter of the words in this frequency group. Surprisingly, as the results indicate, the participants' receptive knowledge began to decline remarkably after about 20 exposures, and continued to fall, with the final group, comprising words met more than fifty times, showing only a 15% gain (7/46), which was exclusively at the receptive level. If the assumption that vocabulary acquisition is incremental in nature (Schmitt, 2000) is to be taken into account, such a decline in receptive knowledge then could be expected and, thus, this result could be interpreted in the following way. The fact that productive knowledge of words comprises also its receptive knowledge, then receptive words could have developed to the productive knowledge after participants encountered the word through 20 exposures or more.

The relationship between incidental learning of words at receptive and productive levels and the number of word encounters in the texts was analysed using a non-parametric Spearman Correlation Coefficient. This was because data in the eight frequency groups was found not to be normally distributed, and thus did not meet the assumption of the parametric Pearson correlation.

To test the relationship between the incidental receptive gain of words from the ER treatment, and the number of word encounters in the texts, Spearman's rho was performed.

Table 6-19: Spearman's correlation between the number of target word repetitions and incidental receptive learning

			<i>Percentage of receptive gain</i>
Spearman's rho	Number of repetitions	Correlation Coefficient	0.310
		Sig. (2-tailed)	0.456
		N	8

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

The analysis in Table 6-19 reveals that, for the eight repetition categories tested, there was no significant relationship between the number of times a word was repeated in the texts and their acquisition at the receptive level ($r=0.310$, $p>0.05$). Therefore it seems that the receptive gain percentage, did not consistently increase with the increase in the number of word exposures.

6.4.2 Productive knowledge improvements

Table 6-20: Productive gain percentages according to each frequency group

Number of occurrences	Improvement	Number of words	%
2-5 (164)	0 → 3	0	2.4
	1 → 3	2	
	2 → 3	2	
	Total	4	
6-9 (335)	0 → 3	6	4.5
	1 → 3	3	
	2 → 3	6	
	Total	15	
10-14 (247)	0 → 3	4	7.7
	1 → 3	8	
	2 → 3	7	
	Total	19	
15-19 (113)	0 → 3	9	20.4
	1 → 3	4	
	2 → 3	10	
	Total	23	
20-29 (130)	0 → 3	14	26.9
	1 → 3	16	
	2 → 3	5	
	Total	35	
30-39 (60)	0 → 3	10	38.3
	1 → 3	10	
	2 → 3	3	
	Total	23	
40-49 (57)	0 → 3	13	38.6
	1 → 3	6	
	2 → 3	3	
	Total	22	
50+ (50)	0 → 3	13	54.0
	1 → 3	11	
	2 → 3	3	
	Total	27	
All words (1156)	0 → 3	69	14.5
	1 → 3	60	
	2 → 3	39	
	Total	168	

For testing purposes, words for which the participants gained Score 3 (productive level) in the pre-test were excluded from this analysis (56-1212). This was because these words had proved to be fully known to the participants and thus, this knowledge would not have the possibility to improve in the post-test.

The results from Table 6-20 above reveal further interesting findings. For instance, the opportunity for picking up a word at the productive level of use was less than 10%, even after 10-14 repetitions. However, unlike the decline in

the percentage of receptive gains after 20 repetitions (as shown in Table 6-18), data in Table 6-20 indicate that productive knowledge growth from ER continued to be enhanced with the increase in the number of word repetitions. The tendency towards an improvement in the productive gain percentages started at 2.4% of words met by participants 2-5 times in the stories they read during the treatment, and continued to rise, reaching a peak of 54% of words encountered more than 50 times. Interestingly, there was just a fractional difference in the amount of productive gains in the 30-39 (38.3%) and 40-49 (38.6) repetition groups. This suggests that exposure to 30 to about 50 word repetitions does not significantly affect incidental productive word learning from reading, and the next point for significant improvement at this level of knowledge is likely to come after 50 exposures (54.0%).

Again, a Spearman's correlation coefficient was computed to measure the strength of the association between incidental learning of words at the productive level of use and the number of occurrences the participants encountered in the stories they read. The results are displayed in Table 6-21 below.

Table 6-21: Spearman's correlation between target word repetitions and productive incidental learning

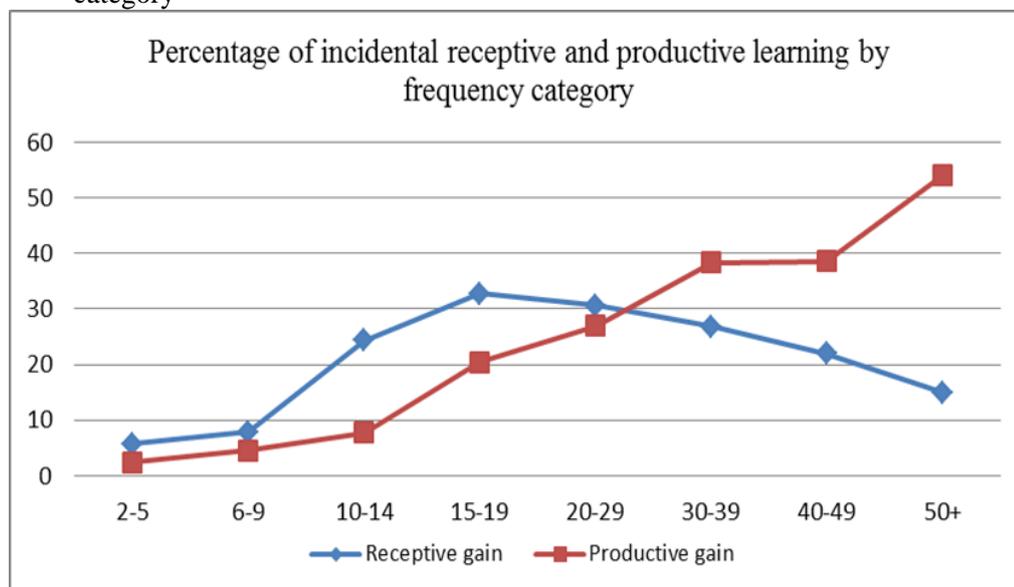
			<i>Percentage of productive gain</i>
Spearman's rho	Number of repetitions	Correlation Coefficient	1.000**
		Sig. (2-tailed)	0.000
		N	8

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

The results indicate that there was a significant positive relationship between productive knowledge growth and the number of word repetitions in the stories ($r = 1.000$, $p < 0.001$). That is, as the number of exposures increased, the percentage of words learnt at the productive level increased.

In addition, an examination of the data, according to the different levels of word gain knowledge also reveals several interesting findings.

Figure 6-1: Percentage of incidental receptive and productive learning by frequency category



As illustrated in Figure 6-1., for the 20-29 exposures frequency group, the chance of learning a word incidentally at a receptive level was slightly higher than the chance of acquiring it at a productive level of mastery. However, the data also revealed a clear tendency towards greater productive gains of words after thirty encounters, in comparison with word receptive level learning, which suffered a remarkable reduction after the same number of exposures. This result suggests that, meeting a word more than thirty times in a text, appears more likely to promote its incidental learning at a productive level rather than merely leading to the recognition of its meaning.

6.4.3 Summary of main findings

Research Question two asks whether the number of times a word is encountered through ER affects its incidental acquisition at a receptive knowledge of meaning level and its productive knowledge at writing use level. Previous vocabulary research on both L1 and L2 learning has established that

the number of exposures to a set of unknown words determines whether the word can be picked up from reading (Saragi et al., 1978; Jenkins et al., 1984; Horst et al., 1998). However, the vast majority of these studies have equated gains in receptive knowledge of meaning with acquisition. Only recently, however, a few researchers have begun to investigate the effects of numbers of repetitions on incidental acquisition at a productive level (Webb, 2007; Chen & Truscott, 2010). The present study is unique, as it examines the relationship between the size of receptive and productive vocabulary gains, at different frequencies of repetition. As stated earlier in this thesis, the receptive knowledge level of a word was confirmed by knowledge of meaning (translation supplementary) whereas, at the productive level of knowledge, the participants were required to demonstrate word meaning knowledge (translation) as well as the ability to use it in a sentence.

In terms of the overall contrast, the effect of the frequency of word occurrences in the graded readers, as a means of enhancing the participants' acquisition of target words at the two levels (receptive and productive) of knowledge reveals several interesting findings. The receptive learning results, as illustrated in Table 6-18, indicate that less than 10% of the words in the 2-5 and 5-9 frequency groups were gained. If word gain is exclusively equated with knowledge of meaning, the results in this table suggest that any word, which is encountered less than 10 times, is very unlikely to be incidentally acquired from extensive reading. The milestone finding is the remarkable three times increase of gain size from 8% after 6-9 meetings, to about 25% after 10-14 exposures, indicating that a greater chance for recognising the meaning of unknown words is possible after 10 encounters. This result seems to be consistent with that of Saragi et al. (1978) who found that an exposure of at least 10 encounters was likely to be needed for picking up word meaning. On the other hand, it does not seem to support other estimates, which have indicated that 6 occurrences in reading are enough for incidental learning of word meaning (Rott, 1999a, p. 11). What led Rott (1999a) to report the need for such a small number of repetitions may be the fact that her highly

controlled study tested the effect of repetition on incidental learning of words immediately after the subjects had met them in six short paragraphs of 4 to 6 sentences. When a test is given directly after reading and the participants are learning words within a very short time and with a modified context, fewer encounters may be required, than when words are met through a lengthier task, such as extensive reading.

Despite the interesting findings in Table 6-18, an ambiguous relationship is suggested between the frequency of encounters and the enhancement of receptive knowledge, particularly after 15-19 occurrences. The correlation test results in Table 6-19 failed to detect any significant relationship between the number of word encounters, and increases in receptive vocabulary size ($p > 0.05$). The gradual improvement in the chances for picking up the meaning of a word was steady up to the point a word was met 20 times in the texts. The surprise, however, is that the number of words learned solely at the receptive level began to decline after this number of word meetings.

A first glance at these findings gives an indication that this study is not consistent with earlier research, which found that word meaning knowledge is strongly enhanced by an increase in the number of in-text word encounters (Saragi et al., 1978; Jenkins et al., 1984; Webb, 2007). When looking at the productive vocabulary results in Table 6-20, however, it can be recognised that receptive vocabulary size is a misleading category. The fact is, that the productive mastery of word knowledge in this study also comprised receptive knowledge. In other words, to use a word productively means that a learner also knows the word receptively. Therefore, it is important to carefully reconsider the apparent drop in receptive gains that occurred after 20 encounters.

In the case of productive gains, there was clearly a steady increase in learning along the frequency scale, starting from 2.4% of words that had been seen 2-5

times in the texts, to half the words being learned (54%) if encountered 50 times or more (see Table 6-20). Therefore, it seems that if there was a noticeable boost of receptive gain of word meanings after 10 occurrences, learning at a productive level of use, for writing requires a greater number of encounters. The percentage of words known at a productive level remained below 10% even after the words had appeared 10-14 times, but increased to 20% in the 15-19 encounters group and continued upwards. As revealed in Table 6-21, the correlation between the number of times the word was met in the ER programme, and the chance for acquiring the word incidentally at this level of use, was very strong ($p < 0.05$). If the greater sophistication of productive knowledge, which also includes receptive knowledge of word meaning, is to be taken into consideration, then the findings should be more comprehensive and indicate that receptive knowledge, alongside the productive knowledge, continued to improve each time the number of receptions increased. This indeed confirms previous findings, which have shown that both receptive and productive vocabulary size improves when the number of encounters with a word is increased (Webb, 2008b).

The results, illustrated in Figure 6-1, reveal further interesting points. Before the productive knowledge gains exceeded the receptive gains (at the 20-29 repetitions category), the participants demonstrated better receptive knowledge (5.7%, 8.0%, 24.3%, 32%) than productive (2.4%, 4.5%, 7.7%, 20.4%), for all of the four, initial frequency groups (i.e. 2-5, 6-9, 10-14, 15-19). However, productive knowledge significantly outweighed receptive knowledge after 20 encounters. These results suggest that, lexical knowledge starts at a basic level of recognition of word meaning, and any further encounters will result in word knowledge development to the productive level of use. This conclusion supports the theory of Nagy and his colleagues (Nagy & Herman, 1985; Nagy et al., 1985; Nagy et al., 1987; Nagy & Herman, 1987), which argues that incidental learning of words from context is a gradual process. Therefore, a possible reason to explain why receptive knowledge gradually declined after 20 encounters, in favour of the productive knowledge, is perhaps the fact that the

chance for the word to move to the productive level was enhanced every time the participants met the word in another context. This seems in agreement with findings from vocabulary acquisition research, which claims that receptive word knowledge gains precede a productive level of mastery, which represents a more advanced stage of acquisition (Aitchison, 1987; Clark, 1993; Melka, 1997).

With respect to the evaluation of the results obtained at each of the two levels (receptive & productive) of knowledge, several conclusions can be drawn. Participants gained a reasonable amount of receptive knowledge of words which they had encountered 10 times in the texts they had read. However, their gain in productive word mastery seemed notably lower after this number of repetitions. This indicates that learning a word at the productive level of use requires a larger number of exposures in ER materials than 10.

Interestingly, however, it was found that 20+ exposures seems to be the borderline where the receptive knowledge of words begins to move to the productive level of use. Any encounter with a word at this number of repetitions, or more, provided a good opportunity for both word knowledge aspects to be enhanced. The fact that word knowledge involves both receptive and productive levels of mastery (Nation, 2001), makes it logical to argue that at least 20 encounters are needed for the word to be incidentally learned from extensive reading. This number confirms the findings of Waring and Takaki (2003) who, using three test formats, found that 20 encounters were required for incidental learning of a word. The number suggested here is obviously larger than has been suggested by many earlier studies. These estimated differences seem largely attributable to type of knowledge measurement used. Unlike early studies which examined the effect of repetition on receptive vocabulary size, using such measures as multiple choice tests, matching L1 words with L2 words, providing L2 meaning or L1 translation, the current

study and the study by Waring and Takaki used measurement approaches with higher levels of sensitivity.

Overall, the results support previous findings (Webb, 2008b), by showing that that receptive, as well as productive, knowledge of words is affected by the number of times a word is encountered in reading. Nevertheless, when reporting the findings of the present investigation, the frequency of exposure is shown to be only one of the factors that can affect incidental word learning from reading. Other factors, such as learners' proficiency levels, the context in which the word occurs, morphology, salience of the word (Pigada & Schmitt, 2006), and the fact that some words from the low frequency list may be easier for the participants to pick up, may also influence the learning of words from extensive reading. While some aspects (namely repetition and proficiency level) have been controlled in this study, others, such as contextual influence, morphology, and word salience are not in the scope of this thesis.

6.5 Research Question 3: How much of the newly learnt vocabulary is retained 9 months after the ER programme?

6.5.1 Comparison between learning retention scores over nine months

Table 6-22 presents the experimental group's vocabulary test mean scores over time: from the pre-test (prior to the intervention), the post-test (one week after the intervention), the second post-test (two weeks after the post-test), and the delayed test (nine-month follow-up).

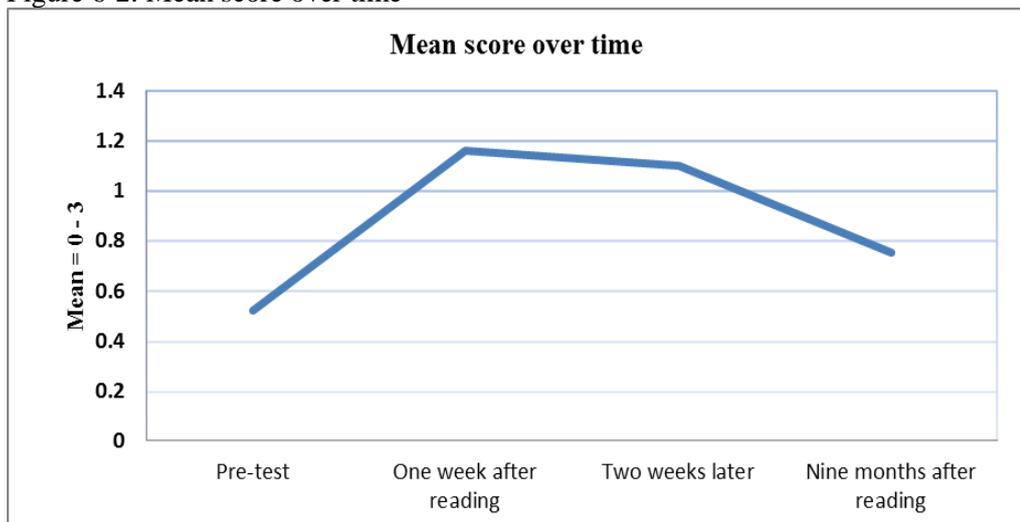
Table 6-22: Descriptive statistics for the vocabulary test scores over time

Test administrations	N	Mean	SD
Pre-test (<i>before reading</i>)	18	0.52	0.14
First post-test (<i>one week after reading</i>)	18	1.16	0.30
Second post-test (<i>two weeks later</i>)	18	1.10	0.31
Delayed-test (<i>nine months after reading</i>)	18	0.75	0.19

Mean score = 0 - 3

In order to compare the mean scores for the vocabulary tests over four administrations, a one-way repeated measures ANOVA test was conducted. This test is mainly performed to compare how a within-subjects experimental group performs in three or more experimental conditions. The ANOVA compares whether the mean of any of the individual experimental conditions differs significantly from the aggregate mean across all of the experimental conditions. The mean and the standard deviations for the four test administrations are presented in Table 6-22, and graphically illustrated in Figure 6-2 below. The data indicates a significant effect for tests over time, $F(3, 51) = 115.84, p = 0.000$.

Figure 6-2: Mean score over time



Although the ANOVA revealed a statistically significant difference between mean scores over a period of nine months, the test does not signify where

exactly this difference occurs. Therefore, to ascertain where the difference between the four test administrations lay, a Bonferroni's post-hoc pairwise¹⁴ comparisons test was conducted. The overall differences in the mean scores between 'the pre-test' and 'the post-test', 'the pre-test' and 'the second post-test', 'the pre-test' and 'the second delayed test', 'the post-test' and 'the second delayed test', 'the second post-test' and 'the second delayed test' were statistically significant ($p=0.000$). However, Bonferroni's pairwise comparison analysis did not reveal any significant difference between 'the post-test' and 'the second-post-test' ($p=0.090$).

The significant difference between the tests revealed by Bonferroni's analysis can be interpreted from the data in Table 6-22 and Figure 6-2 above. The mean scores and the standard deviations indicate that the participants' knowledge of the target words significantly improved when their overall scores in 'the pre-test' are compared with 'the post-test', 'the second post-test' and 'the delayed test' ($p=0.000$). As illustrated in Figure 6-2 however, the mean difference between 'the post-test' and 'the delayed test', and 'the second post-test' and 'the delayed test' ($p=0.000$), indicates a significant decline in the participants' target word knowledge in each of these tests.

So far, the overall mean scores for the target words in the four tests have been compared. Furthermore, since data is available for different levels of word knowledge, elicited at the same point in time, it is possible to compare word knowledge attrition at each of the four levels tested. To track any incidental learning retention, particularly the absolute gain retention (receptive and productive gain), a further analysis, using the percentages of words rated at each of the four levels of the knowledge scale was conducted. A one-way repeated measures ANOVA was performed on each of the four levels of word

¹⁴ Bonferroni's post-hoc pairwise comparisons are commonly performed after significant effects have been found when there are three or more levels to a factor.

knowledge, to determine whether the means differed from each other over the four test administrations. The descriptive data (means and standard deviations) for the four levels of knowledge in the four tests are presented in Table 6-23 and are also graphically illustrated in Figure 6-3 below.

Table 6-23: The mean word percentages by level of knowledge over time

Knowledge Level	Vocabulary test administrations							
	Pre-test		One week after reading		Two weeks later		Nine months after reading	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0* <i>No knowledge</i>	62.01 ^a (742) ^b	5.59	35.02 (420)	7.00	37.41 (440)	9.48	48.28 (580)	8.78
1 <i>Slight knowledge</i>	27.89 (335)	5.27	30.84 (345)	13.02	29.66 (341)	9.10	36.72 (432)	8.07
2 <i>Receptive knowledge</i>	5.80 (79)	3.44	17.51 (227)	7.41	18.19 (237)	7.24	7.01 (95)	4.00
3 <i>Productive</i>	4.30 (56)	4.15	16.62 (220)	10.77	14.74 (194)	9.44	7.98 (105)	5.62

N= 18

*= score mark

a= mean percentage of words rated at scale level

b= the number of words

As Table 6-23 indicates, an analysis of variance through the ANOVA shows that the effect of time on the mean percentages at all four levels of knowledge was significant ($p < 0.05$). The ANOVA outcomes for the levels in Table 6-23 are presented respectively below:

0 (*No knowledge*): $F(1, 17) = 29.952, p=0.000$

1 (*Slight knowledge*): $F(1, 17) = 13.317, p=0.002$

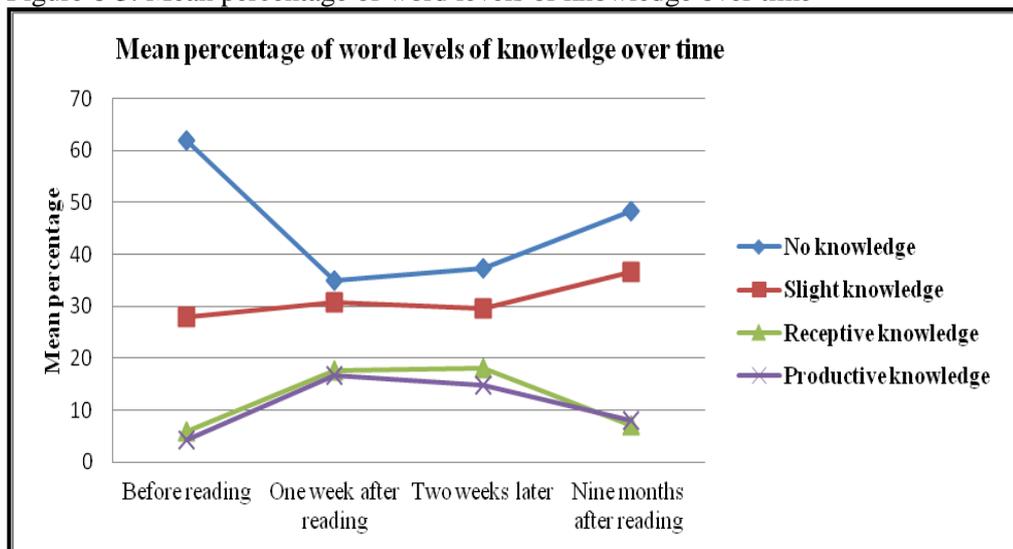
2 (*Receptive knowledge*): $F(1, 17) = 5.403, p=0.033$

3 (*Productive knowledge*): $F(1, 17) = 28.543, p=0.000$

Again, the ANOVA results above demonstrate a significant difference among the four administrations at each level of knowledge, but they do not show how significant the difference was between each two administrations. Hence, to measure the mean difference between the tests, a post-hoc analysis, using

Bonferroni’s pairwise comparison analysis was performed. The difference between ‘the pre-test’ and ‘the post-test’ at all four levels of knowledge was statistically significant ($p=0.000$). This considerable change in the participants’ target word knowledge from the pre to post-test was presented earlier in the analysis of the first research question (see section 6.3.).

Figure 6-3: Mean percentage of word levels of knowledge over time



As seen in Table 6-23 and Figure 6-3. above, the data indicates that there was no obvious change in the participants’ knowledge of the target words, at the four levels, during the second post-test, two weeks after they had taken ‘the post-test’. When this difference was measured through the post-hoc Bonferroni’s analysis, it yielded no significant difference between the means in these two tests at all knowledge levels ($p > 0.05$): no knowledge, $p=1.000$; slight knowledge, $p=1.000$; receptive knowledge, $p=1.000$; productive knowledge, $p=0.107$.

When exploring the difference between ‘the second post-test’ and ‘the delayed test’ however, the means for the four knowledge levels show clear increases and decreases. As illustrated in Figure 6-3, there was an increase in the mean percentage of words participants rated as totally unknown (Score 0) in ‘the delayed test’ compared with ‘the second post-test’. This increase was found to

be significant ($p=0.000$) when a post-hoc Bonferroni's analysis was used. This suggests that many of the target words, which had improved to higher levels of knowledge (slight, receptive, and productive) after the reading programme, may have returned to their original status nine months after the participants had received the ER treatment. A similar mean percentage increase was observed in the number of words participants reported they had seen before (Score 1) in the second post- test and in 'the delayed test'. The Bonferroni pairwise analysis reveals a significant difference between the means in these two tests at this level of knowledge ($p=0.014$).

When looking at the absolute gain, the mean percentage of receptive and productive knowledge shows a clear decline in the participants' knowledge at these two levels in 'the delayed test' in comparison with their performance in 'the post-test' (see Figure 6-3). A post-hoc Bonferroni analysis indicated that the means at these two levels significantly differed ($p=0.000$) in the two administrations.

To investigate whether the participants' pre-intervention knowledge of target words significantly differed in the second delayed test, the difference between the four levels' means was inspected, using a Bonferroni pairwise analysis. The comparison yielded a significant difference at three of the four levels of knowledge ($p=0.000$); that is, as is shown in Table 6-23 and Figure 6-3, the mean percentage of words which were totally unknown (Score 0) in the pre-test had dropped significantly in the second delayed test. Conversely, there was a significant increase in the number of words reported as slightly known (Score 1) by the participants in the delayed test, when the mean is compared with that in the 'pre-test'. Similarly, the increase in productive knowledge from the pre-test to delayed test was also found to be significant ($p=0.000$), when the means were compared by using a Bonferroni pairwise analysis.

Interestingly, however, the Bonferroni pairwise analysis failed to detect any significant difference between the participants' receptive knowledge in 'the pre-test' and 'the delayed test', $p=0.154$. This indicates that nine months after they had the extensive reading, the participants' receptive knowledge had returned to the same level they had before they started the programme.

6.5.2 Absolute gain retention over nine months

Since this study is concerned with the absolute incidental gain of vocabulary at receptive and productive levels, it is worth looking closely at the long-term learning retention of words at these two levels. This was investigated by measuring the participants' receptive and productive knowledge in 'the first post-test' and 'the second delayed test'. The Bonferroni pairwise comparisons above reveal no significant difference between the post-test and the second post-test at all knowledge levels, so it was not considered necessary to include the second post-test in this analysis.

Figure 6-4: Mean percentage of participants' receptive and productive knowledge over time

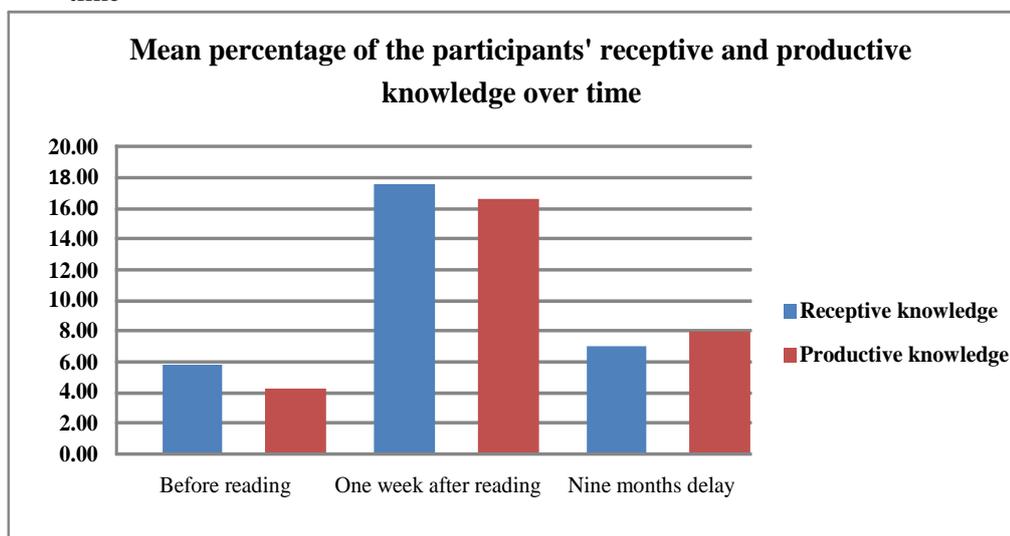


Figure 6-4 shows the means percentages for the participants' receptive and productive knowledge in the pre, post, and the delayed test. The bar graph

clearly shows that the participants' target words knowledge at these two levels grew remarkably after the reading programme. However, this improvement had sharply declined, when the participants were tested on these words nine months later. Their productive knowledge, however, appears to have been better retained in the second delayed test.

Table 6-24: Changes in target word knowledge nine months after the first post-test

T2	T4	Change	Number of words	%
0	→ 0	Stable	274	65.4
0	→ 1	No gain	133	31.7
0	→ 2	Gain	5	1.2
0	→ 3	Gain	7	1.7
1	→ 0	Attrition	200	58.0
1	→ 1	Stable	137	39.7
1	→ 2	Gain	7	2.0
1	→ 3	Gain	1	0.3
2	→ 0	Attrition	67	34.7
2	→ 1	Attrition	90	46.6
2	→ 2	Stable	32	16.6
2	→ 3	Gain	4	2.1
3	→ 0	Attrition	37	22.0
3	→ 1	Attrition	61	36.3
3	→ 2	Attrition	33	19.6
3	→ 3	Stable	37	22.0

T2= one week after reading

T4= nine months after reading

%= percentage of words at each level

Table 6-24 demonstrates the participants' target word knowledge change nine months after the reading programme. The retention of incidental receptive and productive gain from the ER programme is represented in the table by 2→2 and 3→3 respectively. The figures in this table show that, after a nine-month interval, the participants had retained 16.6% (32/ 193) of the words they had

learned at a receptive level from the reading. However, a higher percentage of 22% (37/168) of the words acquired by participants at a productive level of mastery were retained in this delayed test. This means that overall, only about 20% (69/361) of the learning was retained over nine months. Because knowledge decline from productive to receptive knowledge was also considered to be attrition, the 20% portion of retention does not include words that moved backward from 3 to 2 in the delayed test (33 words). This is because many of the words the participants in this study acquired at the productive level, they had known receptively before the reading programme started.

A noticeable point that emerges from these findings is related to where attrition most likely occurs. The results in Table 6-24 show that, the greater the level of the participants' mastery of the word, the less likely it was to retreat to the 'unknown' (score 0) state in the delayed-test. This is clearly shown in the percentages of words in the slight knowledge 1→0 (58%), receptive knowledge 2→0 (35%), and productive knowledge 3→0 (22%) categories. This data suggest that the greater the word knowledge gain, the less likely it was that the word would decay to an unknown state, even for long-term use.

An additional interesting finding, revealed by a cross-tabulation analysis is that only a few words (2.5%) improved to a receptive and productive level of gain (i.e. 0→2, 0→3, 1→2, 1→3, 2→3) during the nine-month interval. This suggests that the participants had a very limited opportunity to encounter the target words after they had finished the reading treatment.

Since the target words were from two different levels of use in English (off-list and 2K words), it was important to check whether the frequency factor affected the mean scores for words from the two categories.

Table 6-25: Mean score differences for off-list and 2K words in the post and delayed test, shown using a Paired-Samples t-test.

	<i>One week after reading</i>		<i>Nine months after reading</i>		<i>t</i>	<i>Sig. (2 tailed)</i>
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>		
<i>Off-list words</i>	1.12	0.29	0.69	0.21	8.27	0.000
<i>2K words</i>	1.22	0.38	0.86	0.27	5.83	0.000

Mean score = 0 – 3

The Paired-Samples t-test data in Table 6-25 reveal an equally significant decline from the post-test mean scores to those in the delayed test, for words in the two frequency categories. The mean for the off-list words in the post-test given to the participants a week after reading (M=1.12, SD=0.29,) decreased to 0.69 (SD= 0.21) in the nine-month delayed test, $t=8.27$, $p=0.000$, whilst mean scores for the 2K words (M=1.22, SD=0.38) in the post-test declined to 0.86 (SD=0.27) in the delayed test, $t= 5.83$, $p=0.000$. This similarity in the mean decrease over time suggests that, despite being from the 2000 most frequent word families used in English, the 2K words selected for this research testing purpose appear not to have been encountered during the nine-month interval.

6.5.3 Summary of main findings

The following summary will serve as an answer to the third major research question posed: how much of the newly learnt vocabulary is retained 9 months after the reading programme? In the present study, two delayed tests were administered to assess the effect of the ER treatment on the participants' long-term incidental vocabulary learning retention. The second post-test (including the 361 words learned out of the total 1212 target words) was taken by the participants two weeks after the post-test, whilst the second delayed test was administered nine months after they had finished the ER treatment.

A few studies have investigated the rate of forgetting over time in incidental word learning from reading (e.g., Waring & Takaki, 2003). However, most of

the research examined earlier in the Literature Review looked at incidental learning in a short-term context, through an immediate post-test. Schmitt (Schmitt, 2010) warns that an immediate post-test should be treated as only showing whether the treatment had any effect on vocabulary knowledge improvement. He stresses that absolute learning must be determined through delayed post-tests. To date, two studies, Waring and Takaki (2003) and Brown et al. (2008) appear to have offered the longest, delayed measurement of incidentally learned vocabulary from L2 extensive reading. It is assumed that, in the ER condition, new words are met repeatedly in various contexts until they are incidentally acquired (Day et al., 1991). This assumption has underpinned research into the human memory, which has indicated that repetition facilitates word retention (Pimsleur, 1967). Due to this belief, the nature of the reading programme design in the current study, where the participants had a great number of word encounters in different contexts, led to the assumption that many words would enter into their long-term receptive and productive use.

To begin with, data from the two week delayed test showed that the participants were able to recall the vast majority of the words they had learned from the ER programme. There was no significant difference between the mean score for the post-test ($M=1.16$) and the second-post-test ($M=1.10$), $p>0.05$. More importantly, learners' absolute retention of these words, at both the receptive and productive levels of knowledge did not significantly decline in the two weeks following the end of the reading programme ($p>0.05$). One might infer from these findings, that if a high level of retention was found three weeks after reading, then the delay in administering the immediate post-test (one week after reading) in the current study would be unlikely to have affected the participants' actual incidental learning gains from the ER programme. This argument is also supported by memory research, which indicates that most forgetting happens relatively quickly after learning (Baddeley, 1990).

These results, however, do not appear to concur with the research findings of Waring and Takaki (2003) who reported a steady decline in their subjects' vocabulary scores for each of the three test formats, after a relatively shorter period of delay (one week after they had the immediate post-test). Although no significant degree for the difference between the immediate post-test and the one week delayed test was noted by Waring and Takaki, it is worth referring to change in the dropped unprompted meaning test (meaning translation test) mean scores, which decreased from 15.3 (SD=3.3) to 11.1 (SD=5.5), as this test closely reflects the test type used in the present study. One possible interpretation for such a striking disparity between the findings in this study and those in Waring and Takaki's research, would be the amount of reading and word exposure in the two studies. Subjects in Waring and Takaki's study were tested on 25 items with appearances varying from one to eighteen times in a single book they read. On the other hand, participants in the present study were tested on a greater number of words (1212), with more repetitions (varying from two to over 100 encounters). The fact that retention of words is more likely to occur when the number of word exposure increases (Pimsleur, 1967) may explain why such significant word knowledge recall occurred in the present research, in contrast with the noticeable rate of forgetting in Waring and Takaki's one week delayed test.

Secondly, the data collected nine months after the participants finished the ER programme revealed different results from those obtained in the first post-test (one week after the participants had finished the ER treatment). Findings indicate that only a small portion of the words the participants had incidentally picked up from reading had a long-term life span. This finding is illustrated in both Table 6-22 and Figure 6-2, which show the sharp increase in the word forgetting rate. At the same acquired level of receptive and productive knowledge, the experimental group had only been able to retain about 20% of the 361 words in the nine month delayed test. This means that the participants were able to recall one word out of every five words they had learned from the extensive reading. This is quite a disappointing learning ratio of return,

considering the effort involved in the two months of extensive reading. In fact, this rate is even lower than that previously reported by Waring and Takaki (2003) and Brown et al. (2008).

It is important, however, to note that the testing instruments in these studies may have determined the learning retention rates. While the present study used an entirely unprompted knowledge testing instrument, subjects in the studies by Waring and Takaki (2003), experienced both prompted and unprompted knowledge testing techniques. Brown et al. (2008, p. 149) themselves acknowledge that “the knowledge [their subjects] needed to complete a translation test seems to be far higher than simply selecting the best answer on an MC test”, when they post-tested their subjects three months after the reading treatment.

The striking rates of forgetting in this study persuaded the researcher to carry out an additional exploration, to determine whether word frequency, as one of the learning factors, had any effect on the learners' ability to retain the words nine months after they had met and incidentally learned them from extensive reading. Overall, it appears that retention was more likely to increase if the number of word repetitions in the texts increased. The figures in Table 6-26 below show the percentage of the target words, in each frequency group, that the participants were able to recall in the nine month delayed test. A conspicuous aspect of these results is that the retention of words at the receptive level seems to have been more consistently affected by the number of exposures than productive knowledge. Consistency in successful productive knowledge retention appeared after the words were encountered more than 20 times in the reading material.

Table 6-26: Word frequency and incidental learning retention rates after nine months

Frequency groups	Retention at receptive knowledge			Retention at productive knowledge			Overall retention		
	No.Words	Retention	%	No.words	Retention	%	No.words	Retention	%
2-5	9	0	0.0	4	1	25.0	13	1	7.7
6-9	25	1	4.0	15	3	20.0	40	4	10.0
10-14	56	6	10.7	19	2	10.5	75	8	10.7
15-19	32	3	9.4	23	2	8.7	55	5	9.1
20-29	38	7	18.4	35	4	11.4	73	11	15.1
30-39	15	5	33.3	23	7	30.4	38	12	31.6
40-49	11	5	45.5	22	7	31.8	33	12	36.4
50+	7	5	71.4	27	11	40.7	34	16	47.1
Total	193	32	16.6	168	37	22.0	361	69	19.1

No.Words: The number of words learnt from the reading programme

Retention: The amount of vocabulary retained nine months after reading

Despite the fluctuation in the relationship between the repetition of the words and the amount of learning retention at the productive level of word mastery, overall, there is a reason to argue that the repetition of words in ER materials can have a positive effect on their retention for long-term use. If the overall results in Table 6-26 are considered, the percentages for recall of words at the two levels of word mastery provide concrete evidence that repetition greatly enhanced the participants' recall of the words after a nine month interval. Participants were able to remember nearly half of the words (47%) they had met more than 50 times.

These interesting findings seem in line with early memory research, conducted by Pimsleur (1967), which indicated that the more often a word is repeated, the longer it is retained over time. This finding refers particularly to research which found that spaced repetition leads to more secure word acquisition than massed repetition (Bloom & Shuell, 1981). A consideration of these findings may therefore raise the question of whether the learning rates reported by studies using an immediate post-test (some of these studies were discussed in the Literature Review) can be considered as evidence of durable incidental vocabulary learning.

One might infer from this research that incidental acquisition is not a matter of one learning event, but that learners need to devote time to rehearsal after words have been acquired. For long-term learning retention of words incidentally acquired from reading, one may suggest that further exposure to words through ER is essential.

One concern about these findings is related to the fact that there was no control over the effects that other language sources may have had on the retention rate of the target words during the interval. However, there was no evidence that other learning took place between the post-test and the administration of the nine month delayed test. It is interesting, therefore, that the results obtained from the experimental group in the delayed test indicate that few, if any, of the words were met elsewhere. The mean scores in Table 6-25 suggest that there was no significant new knowledge of the target words. Instead, there was a clear decline in the mean knowledge score for words from the two frequency categories nine months after the participants had the ER programme. In addition, data in Table 6-24 show that only a small number of words moved from the 'unknown' state to the receptive (5 words) and the productive (7 words) levels of mastery. Likewise, there were only a few cases in which words that the participants had reported as 'had seen before' in the post-test, were receptively (7 words) and productively (1 word) recognised in the delayed test. More interestingly, despite the fact that they had been receptively known in the post-test, only four words improved to the productive level. Overall, a calculation of the number of words that were newly acquired during the nine-month interval revealed a very small proportion, about 2.5% of the words, which had the opportunity of improvement to receptive or productive knowledge (24/957: words of scores 0, 1, 2 in the post-test).

As stated earlier, the 2k target lexical items were from very frequently occurring word families used in the English language, and, as a consequence, it was assumed that the participants would have more opportunities to encounter

these outside the reading treatment materials, in comparison with the off-list words (from a low frequency level of use in English). However, findings confirm that the learning of these words can be solely attributed to the reading material. The significant decline in the mean knowledge of target items from this common set of words ($p < 0.05$), as shown in in Table 6-26, provides strong support to the incidental learning argument made here. If this was the case with the 2K words, then one should be confident of the results obtained from words belonging to the off-list frequency band, as they were statistically less likely to have been met outside the experiment.

Unlike off-list words, it was unexpected that knowledge of words from the 2k category would sharply decline after nine months. The most likely reason for the lack of retention of these target items, which participants were expected to have come across in their university textbooks at the very least, is the fact that the participants had a smaller amount of classroom instruction than normal during the course of the present research and months after. During this period, Libya was in a state of turmoil and educational institutions, like other governmental sectors, were strongly affected by the situation. Schools and universities had to close, particularly in the northern regions of Libya, which witnessed widespread anti-Gadhafi protests, which have been described as a bloody civilian war. The situation in the southern regions of the country was more stable, as no major anti-government protest was reported there. However, there was no regular classroom instruction, particularly at the university education level, as most of this was provided by foreign instructors who had to flee the country, specifically and noticeably after the United Nations approved a resolution authorising a No-Fly Zone over Libya. Most of the syllabus in the English department in the university where this experimental work was undertaken, was taught by teachers from India, Africa, and some Arab states. As many of them had to leave the country, there was not sufficient staff to take over their courses, which consequently led to the cancellation of many classes from the students' timetables. Therefore, the insufficient classroom teaching and the lack of alternative language input due to the very poor EFL context,

may have prevented any possible knowledge improvement of the 2K target words having occurred, when the delayed test was given.

Having considered the data relating to the overall incidental vocabulary learning by the ER participants, which has, in turn, answered research questions 1, 2, and 3, in the following chapter I will present the individualised data from the matched pairs in the experimental and control group, and then focus on the analysis of the individualised reading and incidental word learning by the experimental group.

Chapter 7:

Individual Performance Data

Chapter 7: Individual performance data

7.1 Introduction

This chapter examines research question four, which states: Will individualisation in the selection of ER texts reveal any significant difference in the number of vocabulary items learned, as well as in the learners' perceptions of the ER programme? This chapter seeks to further explore the participants' individual performances, in order to determine whether the individualised design for the ER treatment in this study revealed any significant differences in word acquisition gains, when the amount of reading carried out by each participant is considered.

7.2 Research Question 4: Will individualisation in the selection of ER texts reveal any significant difference in the number of vocabulary items learned as well as in the learners' perception of the ER programme?

The current study was carefully designed, in order to gain an insight into the effects of ER on incidental word learning. The data collection methods primarily involved individualised testing instruments. The most important innovative aspect for measuring new word knowledge was that a complete corpus was created from each target graded reader. The target graded readers had been selected by the participants themselves before the treatment, according to their needs and interests. The 18 ER participants selected and read different texts from a varied range of 40 titles. One important factor present in this study and absent in most other research on L2 incidental vocabulary learning, was that each participant was pre-and post-tested on the same items that had appeared in the books they had personally selected to read. More importantly, incidental learning from the graded books read by individuals in the ER group was confirmed by matching them with control subjects, who did not read the books. The matching samples technique aimed to carefully explore any possible effects that other sources of language input might have had on the learning of the target words. Henceforth, the 18

participants in the experimental group will be referred to as P1 to P18 respectively, while their matching control subjects will be referred to as C1 to C18.

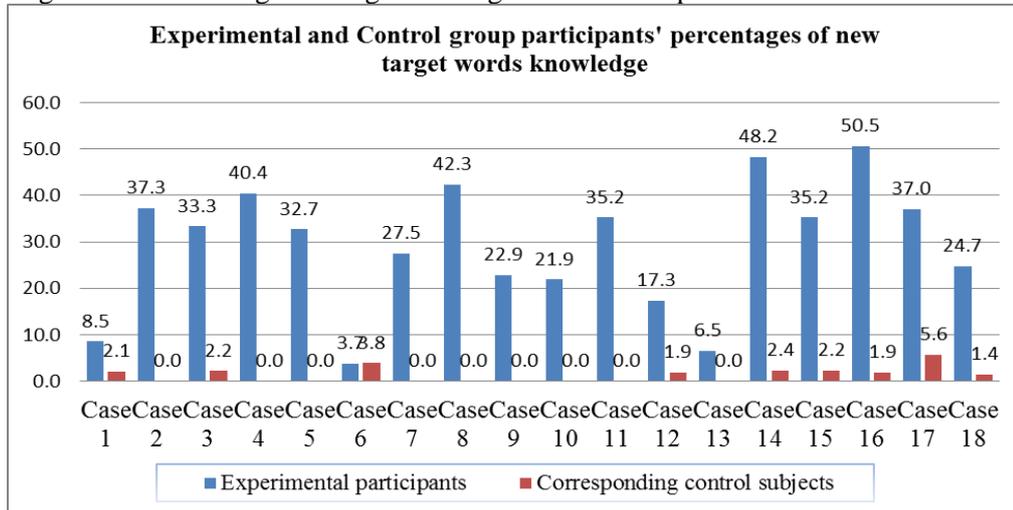
7.2.1 Individual data analysis of experimental vs. control participants' learning

The data presented earlier, in research question one, demonstrated that the difference in the two groups target vocabulary knowledge growth was found to be significant ($P < 0.05$). Overall, participants in the experimental group had learned (at both receptive and productive levels) about a third of the total number of unknown words ($361/1156=31.2\%$) after about two months of extensive reading. In contrast, the control group, who did not participate in the reading treatment, reported new knowledge of a significantly smaller number of words ($16/1164=1.4\%$).

As discussed in the methodology chapter, the design of the present study involved a corresponding control partner for each of the 18 experimental participants. As an initial step in the data analysis for research question 4, it was deemed important to obtain some insight into the vocabulary learning differences between each of the matching pairs in both groups, before making any statement about individualised learning for each experimental participant. Individually examining the significant differences between each member of a matched -pair was not feasible in this analysis, as it was not possible to carry out statistically significant tests on one person at a time. This is because quantitative analysis means that it involves: a quantity of people or units, unlike qualitative analysis: in which a very few observations or situations are described. Therefore, this analysis will include descriptive data of each of the pairs through their individualised learning percentages.

Figure 7-1 below illustrates the percentages of learning demonstrated by each pair (an experimental versus a control participant) after the experimental group was exposed to the ER materials.

Figure 7-1: Percentages of target words gained for each pair after the treatment



An individual learning comparison of the 18 participants in both groups confirms the overall findings from the first research question, which suggested that the experimental group learned significantly more target words than the control group. This is clearly illustrated in the individualised learning percentages of the 18 cases, as shown in Figure 7-1. Apart from the one case (Case 6), where the two matching participants showed a similar rate of acquisition of the target words, the experimental participants in all other 17 cases gained a significantly larger amount of new vocabulary than did their partners in the control group. Furthermore, in comparison with all the participants in the experimental group, who showed various levels of vocabulary knowledge enhancement, the data in Figure 7-1 indicate that half of the control participants (Cases, 2, 4, 5, 7, 8, 9, 10, 11, and 13) did not demonstrate new knowledge for any of the target words, at either the receptive or productive levels.

In order to provide a clearer picture of the level and quantity of word acquisition, a comparison between each matching pair was carried out using their pre/post-

tests scores for the receptive and productive levels. It was considered unnecessary to include partial knowledge gains, as they did not demonstrate the actual level of word acquisition. This comparison data is presented in Table 7-1 below.

Table 7-1: Individualised learning comparison between participants' post-test gains for the experimental group and their corresponding subjects in the control group

Matching pairs	Unknown target words	Receptive Gain				Productive Gain						Total Gain		
		0 → 2		1 → 2		0 → 3		1 → 3		2 → 3		No	%	
		No	%	No	%	No	%	No	%	No	%			
Case1	P1	47	1	2.1	3	6.4	0	0.0	0	0.0	0	0.0	4	8.5%
	C1	47	1	2.1	0	0.0	0	0.0	0	0.0	0	0.0	1	2.1%
Case2	P2	51	6	11.8	1	2.0	5	9.8	5	9.8	2	3.9	19	37.3%
	C2	53	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case3	P3	42	5	11.9	4	9.5	1	2.4	2	4.8	2	4.8	14	33.3%
	C3	46	0	0.0	1	2.2	0	0.0	0	0.0	0	0.0	1	2.2%
Case4	P4	57	8	14.0	4	7.0	5	8.8	4	7.0	2	3.5	23	40.4%
	C4	57	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case5	P5	52	2	3.4	4	6.9	9	15.5	2	3.4	0	0.0	17	32.7%
	C5	54	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case6	P6	27	1	3.7	0	0.0	0	0.0	0	0.0	0	0.0	1	3.7%
	C6	26	0	0.0	1	3.8	0	0.0	0	0.0	0	0.0	1	3.8%
Case7	P7	40	6	15.0	2	5.0	2	5.0	1	2.5	0	0.0	11	27.5%
	C7	40	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case8	P8	71	10	14.1	8	11.3	4	5.6	5	7.0	3	4.2	30	42.3%
	C8	69	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case9	P9	70	4	5.7	3	4.3	0	0.0	6	8.6	3	4.3	16	22.9%
	C9	74	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case10	P10	64	3	4.7	2	3.1	3	4.7	5	7.8	1	1.6	14	21.9%
	C10	58	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case11	P11	71	10	14.1	7	9.9	3	4.2	3	4.2	2	2.8	25	35.2%
	C11	69	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case12	P12	52	5	9.6	1	1.9	1	1.9	1	1.9	1	1.9	9	17.3%
	C12	53	0	0.0	0	0.0	0	0.0	0	0.0	1	1.9	1	1.9%
Case13	P13	77	1	1.3	4	5.2	0	0.0	0	0.0	0	0.0	5	6.5%
	C13	75	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0%
Case14	P14	83	10	12.0	3	3.6	12	14.5	8	9.6	7	8.4	40	48.2%
	C14	85	0	0.0	1	1.2	0	0.0	0	0.0	1	1.2	2	2.4%
Case15	P15	88	14	15.9	7	8.0	3	3.4	3	3.4	4	4.5	31	35.2%
	C15	90	1	1.1	1	1.1	0	0.0	0	0.0	0	0.0	2	2.2%
Case16	P16	99	17	17.2	13	13.1	6	6.1	7	7.1	7	7.1	50	50.5%
	C16	105	0	0.0	1	1.0	0	0.0	1	1.0	0	0.0	2	1.9%
Case17	P17	92	9	9.8	4	4.3	13	14.1	4	4.3	4	4.3	34	37.0%
	C17	89	1	1.1	1	1.1	1	1.1	0	0.0	2	2.2	5	5.6%
Case18	P18	73	4	5.5	7	9.6	2	2.7	4	5.5	1	1.4	18	24.7%
	C18	74	0	0.0	0	0.0	0	0.0	0	0.0	1	1.4	1	1.4%

P= The experimental participant

C= The corresponding control participant

To observe whether any effects from other sources of language input may have had an impact on the incidental learning of words from the ER materials, words which were recognised at the productive level of mastery, by both the experimental and control participants, in the pre-test were excluded in this comparison analysis shown in Table 7-1. This was because the productive level of word use was considered the ultimate level of knowledge; therefore, the participants would have had no chance to gain any further knowledge of these words in the post-test. On a number of occasions, participants recognised some of the target words receptively by just providing their meaning in the pre-test, in these cases words were still included in this analysis as they had opportunity to be transformed to the productive level of use in the time before the post-test.

As discussed in research question one, a small number of the target words had been known productively by both the experimental and control participants when they were first pre-tested. Interestingly, as the data in the table above displays, the pairs' prior knowledge in all cases did not significantly differ. All 18 pairs were mostly post-tested on the same number of words as they had recognised a very small number of the target word in the pre-test. This implies that the comparison between the pairs' learning percentages should be very reliable.

The total gain percentages of target word gains in Table 7-1 show that, with the exception of Case 6, in which the two subjects learned one word of new vocabulary, in all other seventeen cases the experimental participants outperformed their corresponding control subjects. This outperformance was found to be statistically significant, when an overall comparison between the two groups was made, in the analysis for research question one (see section 6.3). While the experimental participants acquired a varied number of the target words, their partners in the control group reported a very small number of newly acquired words. Most noticeably, half of the control group (9 subjects) were found not to have learned any of the target words during the two months in which their corresponding partners took the ER treatment.

Another interesting issue, shown in the post-test data, is that most of the new knowledge the control participants gained was at the receptive level. In other words, participants provided only an L1 equivalent translation (i.e. 0→2, 1→2) for most of the recognised words (9/56.3% words of the total 16 items) without being able to use them in a productive way in sentences. Moreover, 5/31.3% of these 16 words had already been partially known by the control participants at a receptive level in the pre-test. These words were transformed into the productive level of word use (i.e. 2→3).

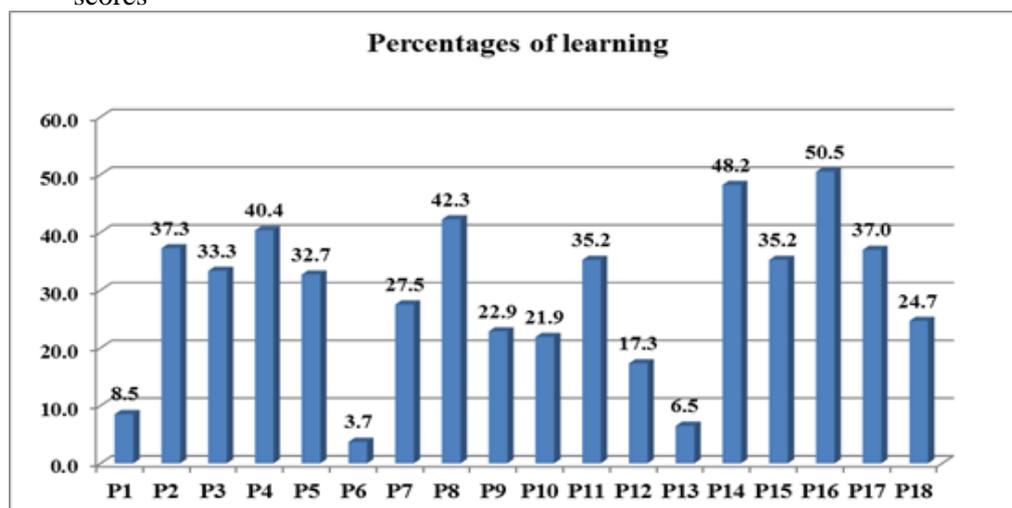
The findings suggest that, in all cases, improved knowledge of the target words by the control group was very limited, in contrast with the new lexical knowledge of their corresponding experimental partners, who appeared to acquire a much larger number of words, both at the receptive and productive levels of word mastery. The findings in Table 7-1 above show that the receptive knowledge growth (i.e. 0→2, 1→2) for the experimental group represented 193 words, learnt by all 18 participants in the group, whereas the control group showed a receptive increase of only 9 words, which were learnt by 7 of the 18 control subjects (i.e. P1, P3, P6, P14, P15, P16, P17). Similarly, only 3 of the 18 participants in the experimental group (i.e. P1, P6, P13) did not demonstrate any improvement in productive word knowledge (i.e. 0→3, 1→3, 2→3) after they had finished the ER programme. The remaining 15 subjects in the experimental group acquired as many as 168 new words at the productive level of use. In contrast, only a small amount of productive improvement (7 words) was gained by 5 of their 18 corresponding partners in the control group. These individualised differences between the pairs in the two groups offer strong evidence to support the overall findings for the first research question, which suggested that the ER groups had learned the target words from the graded readers. It is unclear where the control group picked up the target words, since they were not involved in reading the research material. However, the very small number of words acquired by them, in comparison with their partners in the reading group, who acquired significantly larger numbers of words, does lead to the suggestion that other, English language input, from outside the course, seemed to have a very limited effect on the learning of the

target words. Consequently, it can be confidently claimed that the incidental vocabulary learning demonstrated by all individuals in the experimental group can be accounted for by the ER programme.

7.2.2 Incidental vocabulary growth from individualised reading

This data analysis section will look at the individualised word knowledge growth of the 18 ER group participants. As mentioned above, these participants are referred to as P1 to P18. Data collected through the individualised vocabulary post-test of the group will now be analysed. Factors of participants' attitude toward the ER programme along with their feelings about the programme and the materials in relation to vocabulary learning will also be considered, by means of data elicited from a follow up semi-structured interview with 14 of the experimental groups' participants (i.e. P2, P4, P5, P6, P7, P8, P9, P11, P12, P13, P14, P15, P16, P18). These interviews with the 14 participants were recorded in Arabic and translated into English by the researcher.

Figure 7-2: The experimental groups' post treatment individualised active knowledge scores



As the percentage data demonstrates, differences in the amount of vocabulary gained between the 18 ER participants emerged, despite the fact that they had read books that they, themselves, had selected, which were appropriate for their own

levels of language proficiency. More importantly, the learning ratio amongst this group of learners varied, even though the individualised target words were all selected on the basis that one word in each 1500 words in the books a participant chose to read would be targeted.

By looking at the results in Figure 7-2 very closely, it can be noted that two of the participants (P14 and P16) outperformed the rest of the group. Both of them picked up around half of the items that appeared in their individualised vocabulary test (48.2% and 50.5% respectively). At a slightly lower rate, the data indicate that two other participants (P4 and P8) learned around 40% of their tests items. There were six cases in which participants acquired between 32.7% and 37.3% of the unknown target words (i.e. P5, P3, P11, P15, P17, P2,), and five cases (i.e. P12, P10, P9, P18, P7) where learners picked up between 17.3% and 27.5% of their individualised target words. There were three participants who acquired less than 10% new knowledge of their target words, after they had taken the reading treatment (i.e. P1, P13, and P6).

These individual learning results have revealed very interesting findings. Some of these amounts of lexical acquisition were in line with the high incidental vocabulary learning rates reported in recent studies on word learning (up to 50% learning) from ER material (Waring & Takaki, 2003; Horst, 2005; Pigada & Schmitt, 2006; Brown et al., 2008; Al-Homoud & Schmitt, 2009). For others, rates of learning were as low as those reported in early studies using short texts suggesting that, at best, 10% of the target words would be acquired (e.g. Pitts et al., 1989; Day et al., 1991). It therefore seems important to investigate this further. Part of the reason seems to lie in their actual reading during the programme period. In other words, the number of pages the participants read, appears to be influential in this respect.

As stated earlier in the methodology chapter, the ER participants read texts of different lengths, chosen from 40, varied titles from the Oxford Bookworms

graded books collection. Table 7-2 below contains the levels and numbers of books each of the 18 participants read. It also includes the total number of weeks the participants spent extensively reading their selected books. In terms of the amount of weekly reading carried out by each participant, this information is also summarised in Table 7-2.

Table 7-2: Levels and numbers of participants and their selected books, numbers of ER weeks and amount of reading carried out per week

Participants	Book level	Books selected	Books read	Weeks of reading	Book per week
P1	3	7	7	9	0.8
P2	3	8	8	9	0.9
P3	3	7	7	8	0.9
P4	3	9	9	10	0.9
P5	3	9	9	8	1.1
P6	3	4	4	9	0.4
P7	3	7	6	10	0.6
P8	4	7	7	8	0.9
P9	4	7	7	9	0.8
P10	4	6	6	8	0.8
P11	4/5	*6	6	8	0.8
P12	4	5	5	9	0.6
P13	5	6	5	9	0.6
P14	5	6	6	6	1.0
P15	5	6	6	7	0.9
P16	5	7	7	7	1.0
P17	5	6	6	8	0.8
P18	4/5	**6	6	9	0.7

*.4 books/level 4 and 2 books/ level 5

** .3 books/level 4 and 3 books/level 5

As can be seen from the chart above, the 18 participants read different books from three different language proficiency levels. Apart from two participants (P11 and P18) who selected their books from two levels, all the other participants read books from the same levels. Only three of the participants managed to read one book per experimental week. The reading speeds for the remaining participants were at lower levels and their amount of reading varied from 0.4 to 0.9 of a book, in each week of the programme.

As mentioned earlier in Chapter 5, the participants' individualised tests items were selected from different graded books they had decided to read during the ER

programme. However, because each participant read different texts of different levels and lengths, it was not feasible to create an equal number of lexical items in all individualised tests. This challenge was, therefore, met by targeting one word out of each 1500 in the texts a participant had chosen to read during the ER programme. Table 7-3 shows the number of books selected by each participant and the total number of items included in his/her individualised vocabulary test.

Table 7-3: The amount of participants' extensive reading and number of words included in their individualised tests

Participants	Book level	Number of books selected by each participant	Average word tokens per book	Total word tokens	Number of target words included in each individual participant's test
P1	3	7	10000	70000	47
P2	3	8	10000	80000	53
P3	3	7	10000	70000	47
P4	3	9	10000	90000	60
P5	3	9	10000	90000	60
P6	3	4	10000	40000	27
P7	3	6	10000	60000	40
P8	4	7	16000	112000	75
P9	4	7	16000	112000	75
P10	4	6	16000	96000	64
P11	4(4*) / 5(2)	6	16000/23000	110000	73
P12	4	5	16000	80000	53
P13	5	5	23000	115000	77
P14	5	6	23000	138000	92
P15	5	6	23000	138000	92
P16	5	7	23000	161000	107
P17	5	6	23000	138000	92
P18	4(2) / 5 (4)	6	16000/23000	117000	78

*.The number of books at each level

As shown in Table 7-3, the number of words included in the participants' tests varied between several participants. For example, one participant (P6) was tested on 27 items while P16's test included 107 words. Overall, the table indicates that, the higher the level of books the participants read, the more vocabulary items were added to their tests. For this reason, the final individualised results will not incorporate measurements of vocabulary growth in numbers, as results could be skewed by those who read more and had been tested on a greater number of items. Therefore, it was believed that reporting the pickup ratio of new words from the

various numbers of books the participants had read would make it possible to draw less erroneous conclusions about their incidental word acquisition from ER (see Table 7-4 below). To put it simply, percentage rates measurement served to provide equal learning opportunities for picking up new words despite the fact that the number of texts read and words tested by participants varied significantly. Considering this measurement method, we should be confident that the difference in the participants' amount of reading would have had no great impact on the overall results for rates of individualised word acquisition reported in Table 7-4 below.

During the reading programme, the participants were required to record the number of pages they had actually read in each selected book. This data was collected through the individual book report forms, which were filled out after a participant had finished an entire text. Data collected from this report, along with the participants' individualised pre-post- vocabulary tests scores, were used in this analysis. Therefore, the comparison between the participants' individualised rates of incidental word acquisition was based on what they reported they had actually read. The data in Table 7-4 (below) include information about each of the 18 participants' reading, and their overall incidental word gain percentages.

Table 7-4: Experimental participants and their individualised reading and incidental vocabulary acquisition rates.

Participants	Number of books	Total words	Total actual read words	Actual read words%	Number of target words	Number of unknown words	Number of words learnt	Percent of learning
P1	7	*70000	52750	75.4	47	**47	4	8.5%
P2	8	80000	80000	100.0	53	51	19	37.3%
P3	7	70000	70000	100.0	47	42	14	33.3%
P4	9	90000	82500	91.7	60	57	23	40.4%
P5	9	90000	67750	75.3	60	52	17	32.7%
P6	4	40000	26064	65.0	27	27	1	3.7%
P7	6	60000	51750	86.3	40	40	11	27.5%
P8	7	112000	112000	100.0	75	71	30	42.3%
P9	7	112000	99500	88.8	75	70	16	22.9%
P10	6	96000	93250	97.1	64	64	14	21.9%
P11	6	110000	110000	100.0	73	71	25	35.2%
P12	5	80000	74750	93.4	53	52	9	17.3%
P13	5	115000	49500	43.0	77	77	5	6.5%
P14	6	138000	138000	100.0	92	83	40	48.2%
P15	6	138000	138000	100.0	92	88	31	35.2%
P16	7	161000	161000	100.0	107	99	50	50.5%
P17	6	138000	133250	96.6	92	92	34	37.0%
P18	6	117000	117000	100.0	78	73	18	24.7%

*.The total word tokens in the selected books (average 250 words per page)

**..The numbers of words were not known at the productive level by participants in the pre-test

Table 7-3 shows that the 18 participants selected to read different numbers of titles (varying from 4 to 9), but, the data in Table 7-4 show that not all of the selected books were completed by all of the participants. As the data in the 'Actual read words percentage' column demonstrate, less than half of the group (8 out of 18 participants) read the whole number of texts they had chosen for the ER programme (i.e. P2, P3, P8, P11, P14, P15, P16, P18). By taking a further look at the data, it seems that the incidental vocabulary gain percentages for the 18 learners appears to have been greatly affected by the actual amount of reading they completed in this study.

Prior to factor analysing the results, the hypothesis was that participants who read a larger number of texts would gain greater amount of vocabulary than those who read fewer books. In order to determine if there was any significant relationship between reading different amounts of various texts and the participants' incidental vocabulary acquisition, a correlation test was performed. Prior to that, a test of the normality of the data distribution was carried out. The result indicates that the data in Table 7-4 is non-parametric. Thus, a series of Spearman rank-order correlations were conducted. Since the 18 participants read texts of different lengths, it is more precise and safer to deal with percentages and not the numbers of actual words taken from the total number of selected texts.

Table 7-5: Spearman's correlation between the percentage of actual read texts, reading per week, and the percentage of incidental vocabulary gain

			Percentages of learning
Spearman's rho	Actual read words percentage	Correlation Coefficient	0.693**
		Sig. (2-tailed)	0.001
		N	18
	Book per week	Correlation Coefficient	0.725**
		Sig. (2-tailed)	0.000
		N	18

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

The correlation Spearman's rho test measured the relation between '*percentage of actual read words*' and '*Percentage of learning*'. Not surprisingly, the results in Table 7-5 indicate that the relationship between the two variables was significantly positive, as the Spearman's correlation coefficient r is 0.693 ($p < 0.05$).

Likewise, when the relationship between the amount of reading the participants performed in each experimental week and the incidental vocabulary learning from reading was tested, Spearman's rho revealed a highly significant correlation between the two variables ($r = 0.725$, $p < 0.05$). These findings imply that not only

did the amount of reading the participant carried out enhance their opportunity to pick up new words, but the constancy of reading was also a crucial factor.

The ten participants who did not succeed in reading an entire text from their selected books indicated that they had skipped various numbers of pages. There was one case where one participant read less than 50% of texts she selected (i.e. P13). The remaining nine participants (i.e. P1, P5, P6, P7, P9, P10, P12, P17, and P18) read various, different amounts (from 65.0% to 97.1%) of their books. Some interesting restraints that could possibly have handicapped these participants from completing their reading of these texts during the programme time emerged from the qualitative data and one-on-one interviews with some of the participants (14/18) after they had accomplished their ER treatment further revealed several interesting findings.

One of the responses given, suggested that the participants' attitude towards the reading materials was an important factor that could determine the extent to which the reader continued to read the whole text provided.

... I felt bored with some of the stories, as they were repeating the action, so I did not want to continue reading them. (P6)

Similarly, the simplicity, as well as the length, of some the stories appeared to cause others to have a similar lack of interest in reading:

Some of the books I selected were written in very simple language and were very easy to understand. I did not need to read the whole book to understand what the story was talking about.... (P13)

I selected seven titles from the level four. I thought I would be able to read them all during two months. However, I found it very hard to do that. I consider myself a slower reader... the first book took more than 10 days to finish, so I tried not to spend too much time on the remaining 6 books by leaving out some lines or pages when I felt they were unimportant.... (P9)

Another external factor also seemed to have an impact on the reading practice in this study. Interview data indicated that, for some of the participants, their ability to attend the reading programme on a daily basis was strongly affected by the turmoil, caused by the Libyan revolt at the time the experiment was running.

You know we were not allowed to borrow the books and the problem was that I was not able to come to the university everyday... there was a shortage in car petrol and like many other students, I found it difficult to walk all the way long just to read my books. (P11)

One major problem with the course is the transportation ... I was not able to have extensive reading sessions every day because my father was not able to take me to the university due to the petrol shortage.... (P7)

I wish we had been allowed to borrow the books. It was very difficult to attend to the programme from Saturday to Thursday... taking books home would have allowed us more time for reading the books we selected..... (P5)

In viewing these respondents' comments and their performance in the ER programme, one interesting observation is that, reading practice for some of the participants was hindered by factors such as issues with gaining the materials and the implementation of the course. Therefore, they did not read the entire text of the books they selected to read in this study. As a result, their incidental learning of the words that appeared repeatedly in their chosen titles was also affected by the possible decline in the number of encounters they had with the target words.

The strong relationship between amount of reading practice and incidental vocabulary acquisition is revealed by the results of a Spearman's correlation test in Table 7-5. It can also clearly be seen in the data in Table 7-4 (specifically, '*Actual read words percentage*' and '*Percentage of learning*' columns. With one exception (P17), the highest scores (33.3% and above) were gained by those who had reported reading the entire number texts they had chosen for the programme. In contrast, the two participants (P6 and P13) who had the lowest learning percentage scores (3.7% and 6.5% respectively) were those who had read the

smallest percentages of their books. A further look at data in both Table 7-2 and Table 7-4 above, gives an explanation as to why the relationship between consistency of reading practice and learning of words from the graded materials existed. What probably made P14 and P16 acquire a greater number of words (about 50% of their target words) than those who similarly managed to read all of their selected books, is the fact that, these two participants read more consistently during the treatment. The data in Table 7-2 reveals that these two participants were amongst three participants who managed to read one book per experimental week (P14, P16, P5). However, the possible reason why P5 learned fewer words, despite being engaged in the same rate of reading per week, is probably the fact that she had only read 75.3% of the texts she had chosen prior to the treatment.

The effects of irregular reading on the participants' incidental vocabulary learning were also found elsewhere in the study. Although they were able to fully read their books, some participants (i.e., P2, P3, P8, P11, P15, P18) failed to reach the 50% learning rate. One likely interpretation for this fact, as data in Table 7-2 indicates, is that these participants were less consistent in their weekly reading than those who learned half of their target words. They all read between 0.6 and 0.9 of a book per week.

The participants in this study were asked to select, from three pre-piloted levels of simplified graded book titles that suited their level of language proficiency. Their choices were also checked later by level tests provided by the books' publisher. In the follow-up reading survey, the participants were required to rate their comprehension percentage after fully completing each book. Data in Table 7-6 shows the comprehension percentages for the stories each of the 18 participants read, and the overall mean percentage of story comprehension in total.

Table 7-6: Comprehension rates of different graded stories read by the experimental participants

Participants	Story 1	Story 2	Story 3	Story 4	Story 5	Story 6	Story 7	Story 8	Story 9	Overall mean
P1	100%	95%	95%	95%	100%	95%	100%	--	--	97%
P2	90%	90%	100%	85%	100%	95%	100%	95%	--	94%
P3	95%	90%	100%	85%	100%	95%	100%	--	--	95%
P4	90%	90%	100%	100%	80%	100%	100%	95%	90%	94%
P5	100%	100%	95%	100%	100%	100%	95%	95%	100%	98%
P6	30%	60%	50%	75%	--	--	--	--	--	54%
P7	95%	95%	100%	100%	95%	100%		--	--	98%
P8	95%	95%	80%	100%	90%	95%	80%	--	--	91%
P9	75%	80%	75%	90%	90%	95%	95%	--	--	86%
P10	100%	100%	100%	95%	100%	100%	--	--	--	99%
P11	100%	95%	90%	100%	95%	100%	--	--	--	97%
P12	90%	95%	100%	100%	100%	--	--	--	--	97%
P13	95%	95%	95%	95%	95%	--	--	--	--	95%
P14	100%	90%	90%	100%	90%	100%	--	--	--	95%
P15	90%	95%	100%	95%	100%	95%	--	--	--	96%
P16	100%	100%	95%	90%	100%	100%	100%	--	--	98%
P17	100%	100%	95%	100%	100%	95%	--	--	--	98%
P18	90%	95%	90%	85%	90%	90%	--	--	--	90%

As expected, Table 7-6 shows that many of the participants gained high levels of comprehension. As can be seen, 12 of them rated as much as 95% or above comprehension for the texts they read (i.e. P1, P3, P7, P10, P11, P12, P13, P14, P15, P16, P17). There were only two cases (i.e. P6 and P9) of less than 90% comprehension. If 95% coverage, is considered as a necessary minimum ratio for sufficient comprehension of a text (suggested by Laufer, 1989), the reading for most of the participants in this study does not seem to have been hindered by difficult texts.

In order to determine whether there was any relationship between story comprehension and the amount of incidental vocabulary improvement in this study, the individual learning percentages of the 18 participants (see Table 7-4 above) were weighed against their individual overall mean comprehension scores in Table 7-6. The strength of the association between these two variables was

measured through Spearman's correlation test and the result is presented in Table 7-7, below.

Table 7-7: Spearman's correlation between the participants' incidental vocabulary learning percentages and the level of texts comprehension

			<i>Vocabulary learning %</i>
Spearman's rho	Overall mean comprehension	Correlation Coefficient	0.073
		Sig. (2-tailed)	0.774
		N	18

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

The output in Table 7-7, indicates that there was no statistically significant relationship between the learning of words incidentally from ER and the comprehension level of the texts being read ($r=0.073$, $p >0.05$). This means that any general conclusions drawn from this study need to be tempered by the knowledge that incidental vocabulary acquisition and reading comprehension are not always dependent. It seems that even though the learners gained a high level of understanding of the text, incidental acquisition of words may not have taken place.

In the absence of a relationship between the level of reading comprehension of the various books read by the ER group participants, and their learning of the target words from these books, it is obvious that other factors might have had an important bearing on the amount of incidental acquisition of vocabulary gained. Data collected from the interviewees in this study seem to provide some insight into a variety of these factors.

When they were asked how they dealt with the unknown words they encountered while reading, in the absence of dictionary use, they mentioned several possible learning strategies that they used instead, to learn the new words:

I think the books were written in very simple language...I didn't need to use the dictionary or other strategies, as most of the words in the stories were familiar to me. (P13)

Yes, the dictionary was not allowed and we were asked to guess the meaning. I am not good at guessing and I trust the dictionary, so even when I guess word meanings I can't be sure that is the right meaning for it, so I go and check the dictionary again to confirm that. You know this was not possible in reading my books in this programme, so I don't feel that guessing was a helpful means...(P12).

The comments by these two participants led to an interesting observation. Looking back to the data in Table 7-6, it was found that some participants reported high overall reading comprehension rates (P13=99%, P12=97%), but the data displayed in Table 7-4 suggests that their individual incidental word learning was smaller (6.5% and 17.3%, respectively) than other participants in the group, whose comprehension level was also high. For instance, P4, P8, P14, and P16 acquired a greater amount of word knowledge (40.4%, 42.3, 48.2%, and 50.5%, respectively) from reading at almost the same level of text comprehension (P4=94%, P8=91%, P14=95%, P16= 98%). When interviewed, these participants indicated that they all applied three main strategies to deal with unknown words, when reading their selected stories.

The language of the story helped me to guess the meaning of any new word...the good thing in these stories is the glossary provided at the end of the book. It helped me a lot to find the meaning of unknown words (P16).

We were not allowed to use dictionaries in this programme, which was my biggest problem in reading... I am used to using the dictionary when I read in English. In reading the stories, I had to try to guess the meaning from the sentence. I also found the pictures in the stories very helpful for gaining the meaning of some new words in the books (P4).

At the beginning, when we were told that dictionaries were not permitted while reading and we could use guessing work instead, I said, "how come I will read these long stories?" It is the same, our reading class teacher always encourages us to infer word meaning from the text before we try using the dictionary, but I always had

problem to do that... when I started reading these books, I found that the texts were not very difficult to understand and I knew most of the words in these stories. This helped me to guess the meaning of any new words I encountered in the texts (P2)

I found many of the words I did not know in the stories with the definition at the end of the books....guessing also helped me to discover the meaning of unknown words... (P14)

I think most of the new words were provided with their meaning at the end of the stories.... In reading the first book I guessed the new words, but after one of the students in the programme told me that some words were listed with definitions at the end of the books, I started to look up any new word in these lists and then guessed if the word was not in the list..... (P15)

I guessed the meaning from the sentence. Some pictures helped me to understand the texts. These were the only two methods available at the time of reading. (P6)

The most common strategies indicated by the 14 participants who were interviewed after they finished the reading programme are presented in Table 7-8 below.

Table 7-8: Learners' reported vocabulary learning strategies for discovering the meaning of the new words they met in the graded books

Participants	Vocabulary learning strategies					Incidental vocabulary gain
	Guess word meaning from context	Use book glossary	Analyse part of speech,	Analyse available pictures	Analyse affixes and roots	
P2	Yes	Yes	–	–	–	37.3%
P4	Yes	Yes	–	Yes	–	40.4%
P5	Yes	Yes	–	–	–	32.7%
P6	Yes	–	–	Yes	–	3.7%
P7	Yes	–	Yes	–	–	27.5%
P8	Yes	Yes	–	–	–	42.3%
P9	Yes	–	–	–	–	22.9%
P11	Yes	Yes	–	–	Yes	35.2%
P12	Yes	–	–	–	–	17.3%
P13	–	–	–	–	–	6.5%
P14	Yes	Yes	–	–	–	48.2%
P15	Yes	Yes	–	–	–	35.2%
P16	Yes	Yes	–	–	–	50.5%
P18	Yes	–	–	–	–	24.7%

This table illustrates that, overall, the most common strategies used by participants, for discovering the meaning of new words were: guessing from textual context (93%) and using the book's glossary (57%). Although the strategies of analysing parts of speech, analysing available pictures and analysing affixes and roots were also applied, they were found to be limited to only very a small number of participants, 7%, 14% and 7% respectively. Therefore, in order to explore whether there was any relationship between strategy use in discovering the meaning of unknown words while reading and the rates of incidental vocabulary acquisition by these 14 interviewed participants, these three strategies were excluded from the correlation measurements. The output data from this test is illustrated in Table 7-9 below.

Table 7-9: Spearman's correlation between vocabulary gain and strategies used to determine word meaning

		<i>Guessing from context</i>	<i>Glossary use</i>
Spearman's rho <i>Incidental gain</i>	Correlation Coefficient	.379	.860**
	Sig. (2-tailed)	.182	.000
	N	14	14

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

A Spearman's Rank Order correlation was run to determine the relationship between the 14 participants' reported strategies for discovering the meaning of new words and their percentages of new vocabulary acquisition. The test revealed that the relationship between the learners' guessing strategies and their vocabulary acquisition was not statistically significant ($r_s(14) = .379, p = .182$).

Interestingly, however, the test detected a strong relationship between consistent research of word meaning in the glossary and the percentage of vocabulary the participants acquired incidentally from reading ($r_s(14) = .860, p = .000$). This finding suggests that, even when a reader reads a very large quantity of materials, incidental picking up of new lexical vocabulary is to some extent determined by

strategy use. These data raise a question about the incidental learning in the current study since learning is correlated with a particular strategy, indicating that some noticing was involved when the learners checked word meanings in glossary. However, since this was a self-reported strategy, it is not clear if checking the glossary was the strategy the participants really relied on in learning the target words. Therefore, such findings should be treated with caution. This is particularly important, since the book glossaries included only a small number of newly introduced or uncommon words (see an example in Appendix 10). Therefore, under any circumstances, there was little chance of the majority of words tested in the present study appearing in the book glossaries.

7.2.3 The participants' perceptions of the ER programme

Because ER is considered to be a new approach in the Libyan, EFL classroom, it was thought important to gain a clearer perception of how the participants felt about the programme. Data was gathered from the 14 participants who agreed to be interviewed after they had finished the reading treatment (i.e.P2, P4, P5, P6, P7, P8, P9, P11, P12, P13, P14, P15, P16, and P18). The data collected through one-to-one interviews with each of them reveal several interesting findings. As expected, the participants explained that ER was a new experience for them. In answering the Yes/No question of whether they had experienced ER in English before, all of the 14 interviewees (100%) provided negative responses (No). This was further mentioned by some of them in other instances during the interviews.

This is my first time reading without the teacher observing me and asking me questions about the text. I feel more comfortable reading in this programme than reading in the class. (P15)

Our teacher never provided us with such easy and interesting stories!.... It is a new experience for me and I think reading in this way is more enjoyable and useful than what we read in our classroom. (P 8)

I was very excited It is the first time I have read such long stories throughout a series of sessions (P 4)

The teacher was always responsible for selecting the materials we read, but through this programme, I found that self-selecting texts is more interesting and useful. (P 11)

At the beginning of the course, I was anxious as I never had read this kind of books, but later I was happy. (P 7)

Although, this kind of reading is new for me, I really found it easier and more interesting than our classroom reading. (P 9)

This is my first time I read such interesting stories outside the classroom syllabus. (P 2)

...I think I am not familiar with this kind of reading. (P 12)

The absence of ER in the Libyan context does not seem to be exclusively linked to English materials, as the interview data showed that participants appeared not to read extensively in their own language either. Indeed, only one participant (7%) indicated that she frequently read in Arabic when she had free time.

I frequently read some Arabic books and stories, particularly those written by Taha Hussain such as 'Al-moanoon' (The Sufferers)... I also like reading newspapers and magazines such as 'Al-maraa' (The Woman)... (P 16)

Another interesting finding elicited from the interviews, is that, despite the fact that ER was a novel experience for them, most of the interviewees showed a positive attitude towards it. One of the questions sought to find out how the participants felt about the ER programme. Their comments indicated that about 93% of them (13/14) perceived it in a very positive way

I didn't have much time to read all the books I selected, but from what I managed to read, I found reading in this programme was interesting and useful. (P 5)

This kind of reading is fascinating... I remember that I was very excited reading my books... they were very interesting stories. (P 14)

It was an interesting course...because stories contain lots of conversations. I like that kind of reading! It makes follow the actions easier.... I was eager to know what was coming next... the

reading approach was really exciting and different from what we take in our classroom....(P 7)

I really like I the course. The books and the library we used in this programme were motivating.... (P 9).

I feel it is more fun for me to read for information and pleasure than to read for the exam. Thus, I will try to find and read the same books in the future. (P 16).

This course is really motivating! It makes you like reading... I' ve found myself more comfortable and less stressed with this kind of reading as I read silently on my own (P 11)

Reading was fun and full of action. There were always some new events you got to know through reading. (P 8)

The responses quoted above show that the participants were highly interested in reading the ER materials. Comments by P11, especially, indicate that participants became more motivated to read in English after they had finished the ER programme.

When the participants were asked if they thought the programme was useful, 86% (12/14) of the interviewees believed that the ER programme was beneficial for their language learning, and, 79% (11/14) believed that ER helped them to improve their reading skills.

I think my reading has improved a lot..... Now I can read faster.... (P 11)

Before I took this programme, I lacked the motive to read English texts independently outside the classroom. But now I've built up confidence and I can be more independent in my reading. (P 18)

The programme helped me develop the habit of automatic reading. Since we finished the reading course I've been feeling like I want to read some more similar materials.....I also feel my reading speed has improved. I remember I spent more than 10 days reading the first story, but the remaining 7 stories took less than that amount of time. (P 2)

I feel my English has improved, particularly my reading speed. I think now I can read in English faster than ever. I think have developed my guessing skills when meeting unknown words while reading. (P 15)

Moreover, comments from the participants on the same question revealed that a slightly smaller number of them (71% 10/14) reported that ER had helped them to improve their vocabulary knowledge.

I think I have learnt many new words from these stories. (P 8).

It is interesting that these stories included a very small number of unknown words. This helped me understanding the language and I easily guessed the meaning of the unknown words from the contexts.... (P 2)

I think I learnt new words from the contexts. In many cases an unknown word became familiar to me the next time I saw it in the story. I always tried to guess its meaning in the sentence and I found it was easier with these stories. (P 18)

I now know more words from the story than when I started reading....(P 15)

I learnt many words from the stories. (P 16)

The above data appear to provide a confirmation of the lack of existence of ER as a learning and teaching approach in the Libyan, EFL context. Moreover, the participants' positive attitudes towards the reading programme suggest that, despite its novelty, ER could be a very effective approach for learning, if it were to be introduced to the Libyan EFL university classroom.

7.2.4 Summary of main findings

The analysis of the comparison between the individualised data of the experimental group and their corresponding control participants, who did not take the reading treatment, confirms the findings previously discussed, relating to research question 1. The difference between the matching individuals in both groups was highly significant. In all cases, the experimental participant

outperformed their control partner and none of the control participants' learning exceeded 5% of new knowledge at both receptive and productive levels. One of the main benefits of such a comparison was that a clearer picture about the performance of each pair of matching participants was obtained. This allowed for assurance that the experimental participants outperformed their control partners in all 18 cases. Therefore any growth of the target word knowledge by all 18 ER programme participants can be safely attributed to incidental acquisition from repeated exposure through the graded books.

This research exposed the significance of engaging Libyan EFL learners in extensive reading, particularly the impact of reading different texts on the growth of the learners' lexical knowledge. Despite applying the same procedures for selecting the 18 participants' target words and the acquisition measurement in the present study, the findings showed that the participants' individualised incidental vocabulary acquisition from their self-selected graded books was noticeably varied. Some learners acquired about half of the unknown words during the two months of the extensive reading programme. Other learners, on the other hand, learned as little as 3.7% of the target words. The present study provides empirical evidence that learners' incidental learning of vocabulary from reading may vary due to a number of factors, both external, and related to the learner and these can occur, even if they individually read texts at their level of language proficiency and interest.

The findings presented above indicate that incidental vocabulary learning from ER is not a straightforward task. It appears, from both the quantitative and qualitative analyses presented above, that successful incidental vocabulary learning through ER depends on the presence of several factors. First, the amount of ER the readers engaged in seemed to play an essential role in determining the incidental gain percentages of the target words by the 18 participants in the experimental group. This fact was clearly demonstrated by the correlation test output in Table 7-5. Another major finding was that the majority of these

participants acquired substantial amounts of new lexical knowledge from reading different graded readers. Overall, 78% (14/18) of them learned, at both receptive and productive levels, between 20% and 50% of the words that appeared in their individualised vocabulary tests. This is a particularly encouraging outcome, if compared with the overall gains made by learners, as reported in earlier reading research, are considered here. A number of these studies (e.g., Pitts et al., 1989; Hulstijn, 1992; Dupuy & Krashen, 1993) reported rather smaller incidental acquisition percentages (less than 10%). In these early studies, the subjects read on average as many as 907 words (Hulstijn, 1992) 6700 words (Pitts et al., 1989) and 15 pages of texts (Dupuy & Krashen, 1993). However, due to the larger quantity of texts read by the participants in the present study (varying between about 26,000 and 161,000 words length, see Table 7-4 above), it seems likely that the main reason for these higher rates of incidental vocabulary learning should be attributed to the greater exposure to these words in the ER materials.

The second striking finding in this study is that the amount of reading the participants performed in each experimental week had a great impact on the size of their new incidental vocabulary gains. Those who read more constantly, at an average of one book per week (P14 and P16), had learned significantly larger numbers of target words (about a half of the target words that appeared in their individualised vocabulary tests), than had the other participants in the group. The reason why P14 and P16 outperformed the other participants in the ER group in terms of vocabulary learning seems strongly related to the fact that they managed to read at least one graded book during each week of the experiment. This finding supports the claim that reading one graded reader per week allows sufficient repetitions of words, therefore facilitating the incidental growth of vocabulary (Nation & Wang, 1999). As Table 7-5 above illustrates, a very significant correlation was found between the number of books read per week and the amount of incidental vocabulary gains ($p < 0.05$).

One important aspect of an individual's learning, that was measured in the current research, was that incidental acquisition of words was compared to the actual reading each of the participants performed during the experiment time. The findings showed that many of the participants skipped parts of the texts for several reasons. Because the number of the target word appearances in the stories may have been affected by such text bounce, it can be argued that incidental learning rates could have been enhanced if the learners had read the whole of their selected books.

Thirdly, in terms of reading comprehension, the findings of this study have indicated that the programme was very effective for most of the participants. Both the quantitative and qualitative data in this study demonstrate that the participants, overall, achieved high comprehension rates for their readers. A prior corpus analysis of the graded books (see Chapter 5) illustrated that the vast majority of the running words (85% to 90% coverage) in the graded readers belonged to the 1,000 most frequently word families used in English. It should not be surprising, therefore, that the participants knew a high proportion of the words in their selected stories. If 95% lexical text coverage as the "probabilistic threshold" for adequate comprehension (Nation, 2001) is considered, then 12 out of the 18 participants (66.7%) should have read their stories with ease. An argument concerning the self-report comprehension measurement used in the current study may arise, as participants could have overrated their level of text understanding. However, a vital issue, which seems to validate the participants' rates, is that the selection of their individual levels of readings was based on piloting by reading some pages from each book level, which was later confirmed by a comprehension levels test, provided by the target graded books' publisher. Therefore, such a high level of reading comprehension was expected for most of the participants.

Surprisingly, however, despite the high levels of text comprehension reported by the learners, the correlation between the level of comprehension and incidental vocabulary learning from reading, as the data in Table 7-7 above illustrates, was

not significant. This finding is not in agreement with previous research findings, which have shown that a rich context can help incidental vocabulary learning to a greater extent than a poorer context can (Webb, 2008a). The findings of the present study suggest however, that, even if unknown vocabulary is surrounded by familiar words, this does not necessarily facilitate the incidental learning of new words from reading. It was found that the opportunity for incidental learning of unknown words was not greater, even though learners rated their comprehension at over 95% of the running words in their stories. These results seem to support some research evidence, which has indicated that learners are less likely to notice or remember a new word they encounter if they are caught up in the flow of discourse when they are reading for meaning in a long sequential text (Mondria & Wit-De Boer, 1991; Parry, 1991). Laufer (2003) similarly assumes that when L2 readers understand the overall message of a text, they may not need to pay attention to the precise meaning of individual words. The findings from this study, therefore, suggest that some participants may have read books which were below their language proficiency. Because of the research design purposes, they had to adhere to reading the books they had chosen before the ER programme started, and regardless of the fact that they may have found the books very simple, the participants were not allowed to move up to the next level in order to adjust their level of reading. Thus, encountering words in very simple, rich contexts could have diverted the participants from the lexical level, resulting in a small amount of learning.

One of the major inevitable challenges was the fact that the learners could not change their book reading levels during the treatment period. Typically, ER allows learners to move levels up or down if they feel the text is very easy or too difficult to understand (Day & Bamford, 1998). For the purposes of this study however, this was not feasible. As the qualitative data presented above shows, some of the participants may have read very easy materials. Therefore, being constrained to particular reading levels of books in the current study could have had an effect on the learning ratio. It is possible that many words were encountered in very rich contexts. Thus, learners may have gained the general

meaning of the texts, without the need to know these unfamiliar words. All these facts, therefore, appear to suggest that incidental vocabulary acquisition reported in this study, could have been further improved if the learners had been allowed to freely change their selected levels of books, or perhaps had been asked to select readers from two adjacent levels pre-treatment.

The results presented above, indicate that the types of vocabulary-learning strategies reported to be used by the participants to determine the meaning of unknown words seemed to play an important role in facilitating the incidental acquisition of words from reading. When the participants were asked in the interview about their strategies for finding out the meanings of new words many of them explained that guessing and glossary use were very helpful strategies. Noticeably, despite the absence of dictionary use in this study, in effect, some participants were still relying on it by using the glossary. The correlation test, as shown in Table 7-9 above, clearly shows a strong relationship between the number of acquired items per number of words read by the participants and the type of strategy they said they used to discover the meaning of the unfamiliar words they had encountered. Although these findings could be questioned in the current study as they were based on self-reported strategies, it was found that the participants who reported checking new word meanings in the book glossary while reading learned more words than those who only guessed the meaning from the textual context. In other words, the act of deliberately looking up words helps to commit them to memory, as it focuses the reader on the word's meaning, rather than leading them to swiftly pass over it while reading. These findings support a similar study by Cho and Krashen (1994), who found that the two subjects who consulted a dictionary for the meaning of words acquired more words than the two who just read the story.

By way of comparison, in terms of the current study, one possible reason for glossary use to be more effective in incidental learning of vocabulary, than guessing the meaning from the context, is the operational mechanism of the two

strategies. When using a glossary, accurate definitions of the unknown words are easily accessible to the reader. In that case, L2 learners do not need to process the context and infer the meaning of the unknown words as they do when they apply a guessing strategy. In using the latter, research has shown that learners' guessing of unfamiliar word meanings from the context is not always reliable (Kaivanpanah & Alavi, 2008). Some problems associated with guessing work were found in Laufer's (2003) study. She pointed out that a number of problematic issues can make guessing a less effective strategy in discovering word meanings in reading. These include the fact that (a) some contexts do not provide clues which can facilitate a word meaning inference, (b) contextual clues can be misleading, (c) the reader may ignore clues in the contexts, (d) contextual clues become useless if they occur using words which themselves are new to the learner.

The participants in the present study were only allowed to read the books they had selected before the programme started. Since this prevented them from changing the book level if they found it too easy or too difficult, it is very likely that some of them had to read texts which may have deactivated their guessing skills. According to the IH (Krashen, 1985, 1991, 1997, 2004), language input is more effective if learners read one level beyond their current level of proficiency or what Krashen calls ($i+1$). This is also confirmed by L2 reading comprehension research, which has estimated that understanding of between 95% and 98% of the running words in a text tends to result in effective learning (Laufer, 1989; Hu & Nation, 2000). Therefore, in order to promote more effective guessing, learners need to have great variety of graded materials to choose from. Not only do they need to be encouraged to select and read texts slightly above their level, but they should also be encouraged to change the level if they find it inappropriate for their proficiency level.

Finally, in addition to the statistical findings regarding vocabulary learning, the qualitative data have further revealed valuable insights about the participants' attitudes toward the ER programme. Although the data showed that ER was

considered as a new learning approach in EFL for all participants, it is interesting that the vast majority of them (93%) reported a positive attitude toward it. As mentioned earlier, Libyan university students scarcely ever read EFL materials outside the classroom. The current study, however, found evidence that ER strengthened the participants' motivation to read in English. These important findings indicate that if it were to be introduced to the Libyan universities classroom, ER could receive an enthusiastic welcome from the students.

The noticeable absence of this approach in these classrooms is probably due to the fact that ER is not valued in the Libyan culture. The interview data gathered from these 14 Libyan participants clearly imply only a very small number of them read in their own language. Clarity (2007) points out that L2 learners' experience in their L1 ER may affect their performance in L2 extensive reading. Clarity also suggests that successful implementation of an ER programme in L2 is largely correlated to the learners' interest in the materials and their skill to read in their own language. If this proves true, then any learning conclusion from this study should be taken with care. The findings may represent contexts where L2 learners have a poor habit in extensive reading. It is possible that different results might have been gained if this study were replicated in contexts where learners develop ER in their own languages. Despite the lack of reading habit in both L1 and L2, however, ER appears to have affected the participants' attitude towards reading practice as qualitative data revealed that participants strongly favoured ER over their traditional intensive reading classes. This, in itself, is a positive outcome, which needs to be more widely recognised in this context and which can provide greater impetus for ER to be included more routinely in English learning in Libyan classrooms and beyond.

In addition to the above mentioned factors which may have had an impact on the extent to which each of the participant acquired new knowledge incidentally from the extensive reading, there are other factors to consider which were beyond the scope of this study, such as the level of word difficulty and its impact on

vocabulary test results. . Nevertheless, the findings from the individualized quantitative and qualitative data collected from the participants in the ER group have added to the understanding of incidental learning of vocabulary from extensive reading. Conclusions drawn from these findings will be discussed next. In the following chapter, I will summarise the contribution of my research to knowledge in the field. I will also consider the limitations of the research and make suggestions for further research. Finally, pedagogical implications will be discussed.

Chapter 8:

Conclusions, contributions, limitations, and implications

Chapter 8: Conclusions, contributions, limitations, and implications

8.1 Introduction

This chapter provides conclusions for the present study and presents the contributions it has made to knowledge within the field, beginning with a summary of the major findings which have added to understanding of the role of extensive reading in the incidental learning of vocabulary by Libyan university students. It then considers contributions to knowledge in the field of research design for such studies, which, it is hoped, will serve as a basis for further research studies. The chapter will then go on to consider the limitations to this study, making suggestions for further research, and discussing the pedagogical and empirical implications of the work.

8.2 Contributions to knowledge 1: Major Findings

The two previous chapters have described largely positive findings regarding the incidental growth of vocabulary from ER experienced by adult, Libyan, EFL learners. More specifically, Chapter 6 compared the ER group with the control group and presented the overall positive effect the extensive reading had on the experimental participants' vocabulary development at both receptive and productive levels of use. The chapter also presented findings regarding the relationship between the frequency of encounters with a word and the likelihood of learning, before providing results from the 9-month delayed vocabulary tests. Chapter 7 presented individualised results in relation to the participants' performance in the ER programme, and discussed the amount of vocabulary they had picked up from reading, as well as their perception of the ER programme. This study has made a significant contribution to the state of knowledge about incidental learning of L2 vocabulary from ER in EFL. In the present chapter, the

major findings from the present study are highlighted and the significance of their contributions to incidental vocabulary research will be discussed.

8.2.1 Vocabulary development

This study was carefully designed to gain insight into the effects of ER on vocabulary growth by adult, Libyan, EFL learners. The participants in the ER group read different numbers of interesting stories, which they had selected according to their levels of language competence. The findings offer empirical evidence for the claim that L2 learners can acquire vocabulary through ER. This evidence is also in line with the findings of, for example, Waring and Takaki (2003); Pigada and Schmitt (2006); Al-Homoud (2007). The participants in this study were able to learn almost a third (31%) of their target words at both the receptive and productive levels of knowledge. This means that approximately one in each 3 tested words was incidentally acquired from the ER reading materials. The results have also shown that, overall, the ER group acquired as many as 6 words per day during the ER programme. Although this study was conducted in a very challenging environment (a relatively short programme, with ER being restricted to what the learners had selected before the treatment, and with the disturbance of the Libyan uprising, which happened at the time of this research), the participants displayed a substantial amount of incidental vocabulary acquisition. Thus, one can argue that if ER can demonstrate such favourable results in such a volatile setting, it is reasonable to conclude that it can be a feasible approach for language learning in less challenging contexts.

8.2.2 Receptive and productive growth

This study investigated the relationship between word frequency and incidental receptive and productive vocabulary learning from ER. The results of this study indicate that ER enhanced the incidental growth of vocabulary for adult, Libyan, EFL learners at both receptive and productive levels of knowledge. Some important conclusions about incidental acquisition at these two word knowledge

levels can, therefore, be drawn from this study. Despite the fact that the manner in which learners develop their knowledge from one level to another is not clear, the study suggests that in order to learn a word productively, learners need to encounter it in various contexts, for instance, in a range of ER materials. It appears that learners first acquire the receptive knowledge of a word, and further encounters of it in a variety of different contexts can then facilitate its learning at a productive level of use. This may justify the assumption that vocabulary learning occurs as an incremental process and that word productive knowledge is not only more advanced, but also that it is often learnt later than receptive knowledge (Melka, 1997).

8.2.3 The relationship between frequency of word encounters and learning

This study shows the clear importance of repeated encounters with words in incidental learning and suggests that the more the word is encountered in ER materials, the better the opportunity for receptive and productive gains will be. To give an overview, an analysis of the role that the frequency of word occurrences in the graded readers played in enhancing the acquisition of target words at the two levels (receptive and productive) of knowledge revealed that words were unlikely to be gained at any of these levels of knowledge if they were encountered less than 10 times in the graded materials. This is in line with previous estimates of the number of word encounters needed for incidental receptive gain (Saragi et al., 1978; Webb, 2007). The findings of this study, however, provide a new insight into the number of encounters which may be needed for incidental word acquisition at a productive level. The present study suggests that for incidental receptive gain, L2 learners need to meet to a word at least 10 times. Moreover, a figure of 20 encounters seems to be the borderline at which receptive word knowledge improves to the productive level of use. Therefore, the findings suggest that any encounter with a word at this number of repetitions or above (20+) provides a good opportunity for both word knowledge aspects to be enhanced.

8.2.4 Attrition rates in incidentally learnt vocabulary

While this study has given insights into the amount of vocabulary that can be learned from ER, the findings also suggest that attrition rates of incidentally learnt vocabulary increase over time. The results illustrate that, although a substantial amount of incidental vocabulary was gained by the participants in this study, more than half of the words were lost after 9 months. The nine-month delay test revealed that the participants were able to retain about 20% of the 361 words they acquired from ER materials, but that productive knowledge was more decay-resistant than receptive knowledge. Results also showed that word repetition was an essential facilitating factor in supporting long-term retention. Thus, this study suggests that for long-term retention of words incidentally learned from reading, further exposure to the words through ongoing ER is essential.

8.2.5 Individualised reading and incidental vocabulary growth

An individualised analysis of the ER participant's data demonstrated that there was a strong correlation between the amount of sustained reading each participant practised during the ER programme and their incidental learning of vocabulary. When the participants had read their selected books regularly (about a book per week), they demonstrated a considerably greater amount of vocabulary gain, than others in the group who had engaged in irregular reading activities. This study, however, found no correlation between reading comprehension and incidental word learning. In fact, this study confirmed that, even when participants read with a sufficient level of comprehension, incidental vocabulary gains may not occur from reading. The findings of this study seem to concur with previous research evidence which showed that learners were less likely to notice or remember a new word they encounter if they were caught up in the flow of discourse, when they were reading for meaning in a long sequential text (Mondria & Wit-De Boer, 1991; Parry, 1991).

8.2.6 Learners' attitudes and motivation

The results of this study lend support to the claim that one way to improve learners' attitudes towards reading in a foreign language is through ER (Asraf & Ahmed, 2003). As discussed earlier, reading in English in Libya is restricted to the intensive reading classroom, which often prevents EFL learners from developing a lifelong habit of reading. The qualitative data collected in this study reveals that all the interviewed participants (100%) reported that the ER approach was new for them. Likewise, the data reveal that none of the participants had a habit of reading, either in English or in their own language before the study. The current study, however, found that ER had positive effects on the attitudes and motivation of these participants. After they had finished their books, the vast majority of the interviewed participants (93%) demonstrated positive attitudes towards the reading programme and its activities. Qualitative data have also revealed that participants had improved their motivation for reading in English after the ER programme.

8.3 Contributions to knowledge 2: Innovation in research design

The current study has offered several contributions to the field of incidental vocabulary research and pedagogy that need to be highlighted.

8.3.1 Individualised assessment

The findings from my analysis have answered my research questions and have helped to achieve the study's overall goals. In order to explore the extent to which ER can enhance the vocabulary knowledge of Libyan, university, EFL learners, the researcher used an innovative experimental design, based on an individualised assessment of incidental vocabulary learning by Libyan learners. The implementation of this original design has indicated some valuable methods, which can be used to overcome many of the problematic issues highlighted in

previous research within the field, including identifying the words available for learning over the entire ER programme and assessing the extent to which the subjects learnt them. Previous research has examined the relationship between L2 reading and incidental vocabulary growth, but researchers faced the challenge of determining how much word learning occurred when a learner read longer ER texts. In the current study, this issue was carefully tackled via an approach which was methodologically different from earlier studies. This methodology provides several solutions to previous problems, which may encourage other researchers to adopt the present experimental design.

First, although this study was based on an experimental design initiated by Horst (2005), it was developed in ways that had not been used previously. The new complex procedure refined and developed the methods for selecting the same lexical items for each individualised pre and post-test, which Horst's study failed to offer. In particular, the research started with a corpus analysis which incorporated the entire set of books used in the reading programme. The participants' individualised selection of their ER books several weeks before the reading programme allowed the researcher to use the corpus data of the entire set of books chosen by each of them, in order to select their individualised target words. Therefore, through the pre and post –reading vocabulary tests, it was possible to track the learning of words that were actually encountered in the texts each learner had selected. This important step allowed for more precise procedures for identifying word frequency and for selecting the individualised target words for each participant, before they started reading.

Second, this study adopted a novel design, based on a systematic corpus analysis of the participants' self-selected reading texts. It carefully controlled the measurement of acquisition of a large number of unknown lexical items encountered by the participants through varying numbers of exposures, within different long texts. This innovative strategy added further value to the unique design of the present study, in terms of the way it tackled the challenge of

assessing the participants' actual, individual vocabulary gains from the ER materials.

Another important issue arising from the study is that the corpus analysis method showed that it is feasible for researchers to measure individualised lexical gains by a large number of participants, reading various numbers of ER texts. By adopting a corpus-based research approach, vocabulary researchers will be able to identify, in advance, the target words appearing in the texts each participant chooses to read. By using related software, it is also possible for the researchers to determine the number of times a word will be encountered by a participant in the texts s/he reads. More importantly, the problem of assessing the participants' pre-existing knowledge of the target items from their different texts, which Horst (2005) struggled to solve, is no longer an unachievable goal. The present study indicates that assessment of the participants' pre-reading knowledge of a set of target words appearing in the texts can be solved, in the future, through asking the learners to select their reading texts in advance.

8.3.2 Delayed measurements of incidental vocabulary learning

The present study makes an outstanding contribution in relation to the long-term measurement of incidental vocabulary acquisition from reading. Most of the previous reading studies discussed in section 3.3 reported incidental learning rates of words from immediate post-tests given to the subjects immediately after they had finished reading. There is a strong argument, however, which suggests that the learning reported in an immediate post-test does not give a complete picture of actual acquisition, as a considerable amount of learning decays overtime (Schmitt, 2010). The only study that has attempted to check vocabulary through a 3 month delayed test was carried out by Waring and Takaki (2003). This study, therefore, provides assessment after a far longer delay, probably the longest so far, and therefore, draws a clearer picture of the retention rate nine months after learning from ER occurs.

8.3.3 New research in the Libyan EFL context

A further innovation within this study is the context in which it was carried out. As far as I am aware, there have been no other studies of the effects of ER on incidental vocabulary learning on a sample of adult, Libyan EFL learners, in a university setting. This study, therefore, contributes not only to our understanding of how English vocabulary can be learnt from a programme of extensive reading in Libyan higher education, but also provides evidence of the robustness of this approach in an unusually challenging set of circumstances.

8.4 Limitations of the study

Although I have made many attempts to ensure the quality of the research, there are several inevitable limitations related to different aspects of this complex, experimental design, which should be considered before its empirical and pedagogical implications are addressed.

First, as a case study conducted with a relatively small number of Libyan EFL learners, in one university, its findings are limited in their generalisability to other contexts, within and outside Libya. Although one expects other Libyan EFL learners involved in tertiary education to experience very similar, or virtually identical situations to those at the target university, this experimental research would have been more informative, and the findings would have been more generalisable if the study had included students from the other seven universities. Involving more participants from the other universities would have allowed me to expand the scope of the findings. Moreover, with such an extension, the increase in the number of participants in the reading programme would have allowed me to draw quantitatively stronger conclusions.

However, two major justifications for not including a larger population of learners in this research can be mentioned here. The first reason is related to the complexity of the research design of the present study. The vocabulary learning

assessment was largely based on the corpus analysis of the texts read by the participants and it was therefore not possible to include more participants, as this would have required more time for scanning and corpus work, which was beyond the time scope of the present study. The second reason is related to the critical situation in Libya at the time the present study was conducted. This investigation targeted students in one of eight universities in Libya, which was located in the southern region of the country. Although this part of Libya was, to some extent, less affected by the anti-Gadhafi uprising than other parts of the county, the situation still led to some issues. Fortunately, I managed, with some difficulties, to carry out this experimental research, following the stages discussed in the methodology chapter. Because violent protests spread to all cities in the other regions of the country, however, replication of this study in other universities was impossible.

Second, as referred to in the Methodology Chapter, the ER in this study was tightly controlled. Because of the study design, the participants were asked to adhere to the titles of the graded books they had chosen for the programme. This may have prevented some of them from changing from books which were not interesting or did not suit their language proficiency. For the same reasons, participants were not allowed to borrow or take the ER books outside of the programme library. Additionally, the ER programme was interrupted by the situation in Libya, which affected the sustainability of the learners' reading and the number of texts they were able to read during the weeks of the experiment. These constraints competed with the ER principles of reading a large number of interesting texts, in one's free time. If more typical conditions of ER had been provided, the results of the study might have been different from those obtained.

Third, the initial criteria for choosing the target words in the current study, used whilst piloting, was based on selecting one word in each 500 words the subjects read. However, the piloting study revealed that this would be problematic for the main study participants, since they would be reading much longer texts over a

longer period of time. Therefore, it was important not to overburden participants with longer lists of test items, so the original target number was altered to one word in each 1500 in the texts. This means that many other words which the subjects may have actually learned incidentally from the graded books were beyond the scope of the measurements used.

Fourth, the present study assessed the incidental vocabulary growth of students learning English in an EFL context, where the language input outside the classroom was very limited. Based on the performance of the control group in the pre-and post-tests, the study makes the claim that the experimental group's learning resulted solely from the words met in the graded books and not elsewhere. The EFL context seems to have had little effect on the findings of this study, a result which was concluded from studying the participants' learning rates for words belonging to high frequency word families (2K), as they would have had more chances of being encountered in the participants' textbooks, in comparison with off-list words. This might be due to the circumstances of the study, however. During the research period, the learners participating in this study were exposed to less classroom instruction than normal. The chance of meeting 2K words might, therefore, be higher in a similar context if an ample amount of classroom instruction were provided. Consequently, different results might be obtained from a study which measures word gains from this word family band under normal conditions, in other EFL contexts. Conversely, findings gained from the low-frequency word family category are more robust and likely to be valid for similar EFL contexts, regardless of the difference in the amount of classroom instruction. These words are very rarely used in the English language, so they are not likely to be included in materials designed for second language learners. Hence, these words have very little chance of being met in EFL classroom instruction (Horst, 2005).

Finally, because this study looked at incidental receptive and productive vocabulary acquisition from relatively a short ER programme (9 weeks), it would

be useful to conduct the study for a longer period. This is particularly important for gaining a clearer picture of productive word knowledge development, which, in comparison with receptive knowledge, was found to require a greater number of exposures.

8.5 Suggestions for further research

This experimental research was limited to investigating incidental vocabulary learning from ER by Libyan, EFL students, at one of eight universities belonging to the Ministry of Higher Education. Thus, further studies need to be carried out to widen the scope of this research in relation to the number of universities, target populations, and locations. Such expansion would allow researchers to gain a broader range of results and draw a more complete picture about the incidental learning of words from extensive reading.

In relation to widening the scope of this study, further research could also explore the effects of an ER programme on learners' acquisition of other aspects of word knowledge. Although the study used a 4-point vocabulary knowledge scale (Schmitt & Zimmerman, 2002), the incidental vocabulary measurement included only the two advanced levels of receptive and productive knowledge. Schmitt (2010) argues that repeated exposure to words is likely to lead to the gradual acquisition of other kinds of word knowledge, with perhaps collocational and stylistic knowledge being the ultimate aspects. Thus, it is obvious that there is a need for further research into the relationships between various aspects of vocabulary knowledge. This may require the use of multiple measures of lexis, to better capture a broader range of word knowledge.

The research questions were answered in this study, but at the same time, the findings brought up other interesting issues that deserve special attention in future ER research. For instance, it seems that the act of deliberately looking up words in glossary helps learners commit them to memory, as it focuses them on the word's

meaning, rather than swiftly passing over it while reading. In addition, texts that are too easy seem not to promote incidental vocabulary learning, as the learner may ignore unknown words when they are able to understand the overall meaning of the text. Furthermore, learners are likely to pay more attention to the meaning of words if they consider them relevant to an understanding of the text (Hulstijn, 1993). Therefore, further research needs to consider the question of the difficulty level of the context in which the word is encountered. In other words, harder books may demotivate learners (as may books which are too easy), but the text needs to be at a certain level of difficulty, in order to enable learners to use more effective strategies, involving a deeper focus on meaning, such as word meaning inference from the text (Mondria & Wit-De Boer, 1991; Hulstijn, 1992). Due to this design issue, participants in this study were asked to adhere to the titles and levels they had chosen to read before the reading programme started. For future research, achieving a balance between text difficulty and strategy use could be attained by allowing the learners to move up and down between levels and change texts if they find them either too difficult or easy.

8.6 Pedagogical implications

This study suggests that the effect of ER on incidental vocabulary learning deserves much more attention than it has received to date, particularly in a context such as Libya, where students are unlikely to have developed the habit of reading outside of the classroom, either in their L1 or L2. As in previous research, the current study also concurs that ER can facilitate the incidental growth of L2 learners' vocabulary. Therefore, a number of pedagogical implications are suggested.

Firstly, this study has affected my own method of teaching English in the Libyan university classroom, and will stimulate changes in my future classroom practices of reading and vocabulary teaching. Therefore, I hope to encourage my colleagues, as well as my students, to explore new methods of vocabulary

knowledge enhancement through the ER approach. Findings from this study also have some possible contributions to the development of the methods of teaching, and the reading materials used, particularly in the researched institution, and, in Libyan universities generally. Indeed, the research has provided new knowledge that will pave the way for further research. Libyan learners, like many other learners in similar EFL contexts, are expected to learn an enormous number of words in order to master the English language. ER can be an effective means for that. Research indicates that for language comprehension to occur, learners need to know approximately between 95% and 98% of the running vocabulary in a written discourse (Hu & Nation, 2000). As in many other EFL contexts, the instruction time devoted to English in Libyan classrooms seems insufficient to cover such a large number of words. In ESL contexts, learners gain a benefit from much greater exposure to the English language and culture available outside the classroom. This is, unfortunately, not the case in EFL contexts, where learners have no or little access to such exposure. This study suggests that learners in input-poor EFL contexts can be exposed to greatly enhanced input, if their institutions can make a large supply of books and other print materials available to them. In addition, the benefits of ER can be exploited by integrating this approach into the English curriculum in Libyan tertiary institutions in particular, and in similar EFL contexts elsewhere. If reading is to be exploited for incidental vocabulary learning, it should not be limited to the intensive reading approach and ER should be incorporated into the curriculum. A number of researchers (Hulstijn et al., 1996; Nation, 2001) emphasise that incidental and intentional learning, through extensive and intensive reading respectively, are two complementary approaches to vocabulary learning.

The qualitative data collected from the participants in this study also indicated that ER has, in the past, played no part in the Libyan tertiary curriculum. All the interviewed participants (100%) reported that, not only was this their first exposure to ER activities in English, but also that they had no habit of reading in their own language. This was perhaps due to their cultural and educational background. The overemphasis on the intensive mode in teaching reading in the

Libyan EFL context has left very little room for implementing extensive reading. Undoubtedly, both the absence of this approach, and the lack of ER materials in the Libyan EFL tertiary curriculum, have also had an impact on English teachers' knowledge of the approach. Therefore, this study suggests that in such a context, where ER is totally novel, some groundwork is required for successful implementation of this type of reading programme.

Since ER has been found to be a non-existent approach in Libya, the study indicates that the initial steps of preparation for introducing ER are currently very important. These steps should involve both teachers' professional development and learner training for ER. One possible way to enhance the quality of teachers' knowledge of ER is through the provision of teacher development courses by educational institutions. These courses should aim at improving teachers' knowledge and understanding of ER and its role in teaching English language in general, and improving incidental vocabulary growth in particular. Teachers also need to be provided with the resources they need to develop their skills in teaching extensive reading. Training courses should cover how teachers can successfully prepare and set up ER programmes. They may also include modules on how to work with students in a collaborative way, how to motivate them towards a more independent reading approach, how to bring outside experiences into the classroom to encourage ER and how to promote reading outside the classroom.

As discussed earlier, in the introductory chapter, due to factors relating to the students' cultural background and the education system, Libyan EFL learners lack the habit of ER. Therefore, before setting up any ER programme, teachers need to express the purpose of the approach clearly to the students by holding discussion sessions about the benefits of the ER for language learning. Moreover, teachers should develop their students' awareness of the importance of ER in enlarging their vocabulary knowledge. Paribakht and Wesche (1999, p. 216) comment on "the potential usefulness of building learner awareness of how new vocabulary

knowledge may be accessed through reading and related activities”. Moreover, in EFL contexts where ER is not commonly used, teachers will need to promote their students’ enthusiasm for autonomous learning, by providing them with a variety of interesting reading materials written at varying levels of difficulty. With their teachers’ guidance and encouragement, students could start by reading short, easy books. The greatest benefit of reading easy texts is that students will be able to read faster and with high levels of comprehension, so they can develop an awareness of the difference between their traditional intensive reading experiences (reading in detail with specific learning aims and tasks) and the fluency of ER which involves them in reading long or large quantities of texts for general understanding, with the intention of enjoying the texts (Waring, 2011). This is believed to help students to become more confident and motivated, to sustain the process more successfully and, consequently, to develop the habit of extensive reading.

This study also suggests that ER materials could be an effective source of vocabulary learning outside the classroom, if learners choose and read interesting materials at an appropriate level of proficiency. The findings indicate that, in Libya, and probably other similar EFL contexts where learners seem to be reluctant EFL readers, intensive reading should not be the only approach to teaching and some space should be offered for ER to be integrated within the school curriculum. It is suggested that, instead of involving EFL learners in intensively reading the same book for the whole academic year, which may be demotivating and boring, providing large quantity of books for ER, belonging to various levels and on topics can help to increase the learners’ motivation for reading and build the habit of reading in English

More importantly, the findings show that the incidental acquisition of vocabulary from ER is enhanced when learners utilise certain learning strategies, a fact which has also been confirmed in previous research (Cho & Krashen, 1994). Although the learners were not allowed to use a dictionary to look up unknown

words during this ER programme, learners who reported looking up unknown words in glossaries showed better vocabulary knowledge improvement than those who exclusively guessed new words from their context. Thus, in order to gain the greatest benefit from extensive reading and to build on their incidental learning, learners, perhaps poor readers in particular, may need to be allowed to consult a dictionary for unknown words. However, teachers should explain to them that dictionary use should not constantly interrupt reading and therefore jeopardise the main goal of ER, which is to read large quantities of text for general information. If reading materials are at an appropriate level for the learners' proficiency, however, as previously discussed, such use of glossaries or dictionaries should be minimal and, thus, not interfere with reading fluency.

The corpus analysis of the texts in the current study indicated that many of the unknown words in the graded readers appeared a smaller number of times than is suggested, in this research, to be necessary for receptive and productive knowledge growth (20 times). If graded books are to be used for the purpose of vocabulary growth then a massive amount of such material is needed to build new vocabulary. This carries an important implication for teachers who are willing to introduce extensive reading to their students, i.e., a greater number of books and sufficient time need to be provided for an effective ER programme. This will allow them to choose and read large quantities of texts which, in turn, will help in increasing the number of word encounters they experience.

When it comes to the long term retention of incidental vocabulary, the findings confirmed some concerns. Despite the substantial incidental learning reported in the present study, the participants forgot many of the words they had acquired from the reading materials nine months after they had finished the programme. This finding suggests that the setting up of a one-off ER programme may not be enough to ensure long term gains. For instance, Nation (2001, p. 155) acknowledges that incidental vocabulary acquisition from reading is very 'fragile'. Therefore, in order to retain learning for long-term use, words need to be

quickly reinforced by other meetings. That is, more pedagogically effective incidental word learning would be achieved through further, sustained, extensive reading. Students need to be encouraged to have regular sessions of ER during their EFL education period, or perhaps adopt it as a life-long approach. A suggestion made in the present study, and confirmed in earlier research (Nation & Wang, 1999), is that one graded book per week can result in enough meetings of words to facilitate incidental learning.

8.7 Empirical implications

There is increasing concern about distinguishing the source of learning when actual words are targeted in the texts. A number of studies used non-words as a technique to ensure that there were no effects from outside language input. However, the fact that incidental vocabulary assessment should feature words that are actually encountered in the ER materials (Horst, 2005) should not be undervalued. It is perhaps possible to substitute a small set of target words with non-words when the researcher uses a single text in a study (e.g., Waring & Takaki, 2003), but this is unlikely to be relevant in research incorporating a very large number of target items in various, far longer texts (Al-Homoud, 2007). This issue has been one of the most daunting challenges for research involving large numbers of ER books. Al-Homoud (2007) for example had to use two versions of the VLT to estimate the vocabulary growth of Saudi learners involved in an ER programme.

Recently, research has used very low-frequency word families from the L2 texts learners have read. With the advances in hardware and software capabilities and widespread access to web-based software, vocabulary researchers should have no further difficulties working with target words in large numbers of texts. The design of the current study suggests that corpus analysis is a valuable resource, and helps determine the frequency levels of words encountered in various ER texts. It also indicates that corpus analysis is very important if words from low

bands of frequency are to be used in a study including large numbers of books and participants. For future ER research, the procedures in this study indicate that scanning and creating a huge corpus database from a great number of graded books is possible, provided that sufficient time and funds, as well as helping hands, are available to the researcher.

The delayed tests carried out showed the importance of delayed assessment of incidental word learning from reading. An immediate post-test may be necessary to give insight into incidental learning rate from reading, but researchers should be aware of the fact that attrition occurs in any learning (Schmitt, 2010). The present study showed that the retention rate of the incidental acquisition of words from ER decreased over time. It is hoped that any further research will also obtain long-term decay data, in order to give an accurate picture of the actual learning that has occurred from reading.

Similarly, this study has demonstrated the need for multiple measures of vocabulary gains. The findings show that word knowledge is acquired at different levels and one aspect of knowledge is likely to be gained before another. The use of knowledge scale in this study clearly indicates that learners acquired a word at different levels of use (i.e. receptive and productive levels of mastery). Therefore, it is recommended that “any single measure of it will give only a very minimal impression of the overall lexical knowledge constellation” (Schmitt, 2010, p. 152).

8.8 Personal reflections

As mentioned in Chapter 1, the genesis of the current study can be traced back to when I was teaching adult, EFL learners in Libya. The clear, positive impact of reading on the vocabulary learning of one of my students motivated me to learn about lexical development through ER, hoping to inspire my students to increase their incidental vocabulary learning. Through my experiences, I have found that

my learning can be best communicated to adult, Libyan EFL students through concrete examples from the research that I have personally carried out. This experimental study has been an epic journey, which has developed as I have progressed and has enriched my understanding of this topic. In this research, I was determined to test an intervention, through which I hoped to develop, and add to the methods my students used to increase their EFL vocabulary knowledge and hopefully change the way in which my university department taught reading and vocabulary learning.

Lessons learned:

Since beginning my doctoral studies, I have found the post graduate research environment in the School of Education, University of Nottingham, to be personally and academically stimulating. I began as a novice researcher, with limited knowledge of educational research, however, during my journey I have developed important skills as a researcher, particularly in terms of research methodology and procedures, including the design, collection, analysis and presentation of research data. The research methodology classes that were provided in my first year as a researcher were valuable, not only in fostering these skills, but also in building the critical thinking skills which have been necessary for my doctoral studies. The research knowledge I have developed will be utilised for my future research endeavours and professional development.

During this journey, I have become more acquainted with the diverse literature that has been produced within the field and have consequently widened my knowledge of this research area. This experimental research has helped me to broaden my understanding about both vocabulary learning and research, and I have developed a deeper understanding of some of the key issues, that can affect L2 learners' acquisition of words from reading.

Before embarking on this study, I was of the view that research is a straightforward, linear activity which can be planned and conducted in a set way

and within a fixed period of time. However, the process has presented unexpected challenges. For example, the death of my main supervisor (Dr. Richard Pemberton), after a year of valuable supervision and guidance, as well as the state of unrest in Libya during the time this study was carried out created some difficulties in terms of conducting the research within the planned time frame. As a result, during the intervention, I was left alone to carry out the complex experimental procedures because, at that time, the communication networks in Libya broke down, as a result of the uprising, and there was no means of contacting my supervisors in the UK. I, therefore, have learned to work more independently, which, in itself, was another important gain that I am able to take away with me from this research process.

References

- Aitchison, J. (1987). *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford: Blackwell.
- Al-Homoud, F. (2007). *Vocabulary acquisition via extensive input*. PhD Thesis, University of Nottingham.
- Al-Homoud, F. and Schmitt, N. (2009). Extensive reading in a challenging environment: A comparison of extensive and intensive reading approaches in Saudi Arabia. *Language Teaching Research*, **13**(4): pp. 383-401.
- Al-Issa, A. and Dahan, L. S. (2008). Prior knowledge and writing in the college EFL composition classroom. In Coombe, Cjendli, A. and Davidson, P. (eds.) *Teaching Writing Skills in EFL: Theory, Research and Pedagogy*. Dubai: TESOL Arabia Publications, pp. 17-26.
- Al-Khasawneh, F. (2010). Writing for academic purposes: Problems faced by Arab postgraduate students of the college of business. *ESP World* [Online]. Available from: http://www.esp-world.com/Articles_28/WRITING.pdf [Accessed: 16/3/2014].
- Al-Naiely, S. (2012). *The impact of specialized schools on the quality of the university EFL education*. MA Dissertation, University of Almergib, Libya.
- Alahirsh, H. (2006). *Exploring vocabulary learning strategies in the libyan university EFL classroom*. MA Dissertation, University of Nottingham.
- Aldabbus, S. (2008). *An investigation into the impact of language games on classroom interaction and pupil learning in libyan EFL primary classrooms*. PhD Thesis, Newcastle University.
- Amer, A. (1997). The effect of the teacher's reading aloud on the reading comprehension of EFL students. *ELT Journal*, **51**(1): pp. 43-47.
- Anderson, G. and Arsenault, N. (1998). *Fundamentals of Educational Research*. 2nd ed. Hampshire: The Falmer Press.
- Anderson, J. (2005). *Mechanically Inclined: Building Grammar, Usage, and Style into Writer's Workshop*. Portland, Maine: Stenhouse Publishers.
- Anthony, H. M., Pearson, P. D. and Raphael, T. E. (1993). Reading comprehension: A selected review In Cleary, L. M. and Linn, M. D. (eds.) *Linguistics for Teachers*. New York: McGraw Hill, pp. 282-307.

- Asker, A. (2011). *Future self-guides and language learning engagement of English-major secondary school students in Libya: Understanding the interplay between possible selves and the L2 learning situation*. PhD Thesis, University of Birmingham.
- Asraf, M. R. and Ahmed, I. S. (2003). Promoting English language development and the reading habit among students in rural schools through the Guided Extensive Reading program. *Reading in a Foreign Language*, **15**(2): pp. 83-102.
- Azeroual, D. (2013). Investigating the reading difficulties of Algerian EST students with regard to their general English knowledge. *Arab World English Journal*, **4**(1): pp. 203-212.
- Baddeley, A. D. (1990). *Human Memory: Theory and Practice*. London: Lawrence Erlbaum Associates.
- Bamford, J. (1984). Extensive reading by means of graded readers. *Reading in a Foreign Language*, **2**(2): pp. 218-260.
- Bamford, J. and Day, R. (1997). Extensive reading: What is it? Why bother? *The Language Teacher*, **21**(5): pp. 6-8.
- Bamford, J. and Day, R. (2004). *Extensive Reading Activities for Teaching Language*. Cambridge: Cambridge University Press.
- Barnett, M. A. (1989). *More Than Meets the Eye: Foreign Language Learner Reading Theory and Practice*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Bauer, L. and Nation, P. (1993). Word families. *International Journal of Lexicography*, **6**(4): pp. 253-279.
- Bell, T. (1998). Extensive reading: Why? and how. *The Internet TESL Journal*, **4**(12): pp. 1-6.
- Bensoussan, M. and Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, **7**(1): pp. 15-32.
- Bernhardt, E. B. (1991). *Reading Development in a Second Language : Theoretical, Empirical, and Classroom Perspectives*. Norwood, N.J: Ablex.
- Bernhardt, E. B. (2011). *Understanding Advanced Second-Language Reading*. New York: Routledge.

- Black, I. (2007). Great grooves and good grammar. *The Guardian*. [Online]. April, 10. Available from: <http://www.guardian.co.uk/education/2007/apr/10/internationaleducation/news.tefl> [Accessed: 02/03/ 2011].
- Blaikie, N. (2000). *Designing Social Research*. Cambridge: Polity Press.
- Bloom, K. and Shuell, T. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, **74**: pp. 245-248.
- Brown, R., Waring, R. and Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, **20**(2): pp. 136-163.
- Bryman, A. (2004). *Social Research Methods*. 2nd ed. Oxford: Oxford University Press.
- Bryman, A. (2008). *Social Research Methods*. 3rd ed. Oxford: Oxford University Press.
- Bugaighis, W. (2011). *Libya: Authentic voices from around the world* [Online]. Available from: <https://www.world-affairs.org/wp-content/uploads/2012/03/authentic-voices-libya.pdf> [Accessed: 10/03/2013].
- Carbery, S. and Yoshida, R. (2003). *Graded vs Authentic: Do graded readers take the fun out of reading?* In Proceedings of the 16th EA Educational Conference 2003. English Australia. Melbourne, Australia. Available from: http://www.englishaustralia.com.au/ea_conference03/proceedings/pdf/009F_Carberry_Yoshida.pdf [Accessed: 26/11/2012]
- Carey, S. (1978). The child as word learner. In Halle, M., Bresnan, J. and Miller, G. (eds.) *Linguistic Theory and Psychological Reality*. Cambridge: MA: MIT Press, pp. 264-293.
- Cargan, L. (2007). *Doing Social Research*. Plymouth: Rowman & Littlefield.
- Carrell, P. and Carson, J. (1997). Extensive and intensive reading in an EAP setting. *English for Specific Purposes*, **16**(1): pp. 47-60.
- Carter, R. and McCarthy, M. (1988). *Vocabulary and Language Teaching*. London: Longman.
- Carver, R. P. (1992). Commentary: Effect of prediction activities, prior knowledge, and text type upon amount comprehended: Using reading theory to critique schema theory research. *Reading Research Quarterly*: pp. 165-174.

- Chen, C. and Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, **31**(5): pp. 693-713.
- Cho, K. and Krashen, S. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, **37**(8): pp. 662-667.
- Clarity, M. (2007). An extensive reading program for your ESL classroom. *The Internet TESL Journal* [Online], **8**. Available from: <http://iteslj.org/Techniques/Clarity-ExtensiveReading.html> [Accessed: 29/12/2013].
- Clark, E. (1993). *The Lexicon in Acquisition*. Cambridge: Cambridge University Press.
- Coady, J. (1993). Research on ESL/EFL vocabulary acquisition: Putting it in context. In Huckin, T., Haynes, M. and Coady, J. (eds.) *Second Language Reading and Vocabulary Learning*. Norwood, NJ: Ablex Publishing, pp. 3-23.
- Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In Coady, J. and Huckin, T. (eds.) *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge: Cambridge University Press, pp. 225-237.
- Coady, J. and Huckin, T. (eds.) (1997). *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge: Cambridge University Press.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, **25**(3): pp. 301-315.
- Cobb, T. (2010). *Web Vocabprofile (Version 2.5) [Computer software and manual], an adaptation of Nation & Heatley's (1994) Range* [Online]. Available from: <http://www.lexutor.ca/vp> [Accessed: 27/05/2010].
- Cohen, L., Manion, L. and Morrison, K. (eds.) (2007). *Research Methods in Education*. 6th ed, London: Routledge Falme.
- Cohen, L., Manion, L. and Morrison, K. (2008). *Research Methods in Education*. 7th ed. London: Routledge.
- Cook, V. (1996). *Second Language Learning and Language Teaching*. London: Arnold.
- Cooper, M. (1984). Linguistic competence of practised and unpractised non-native speakers of English. In Alderson, J. A. and Urquhart, A. H. (eds.) *Reading in a Foreign Language*. London: Longman, pp. 122-138.

- Craik, F. and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**(3): pp. 268-294.
- Crotty, M. (1998). *The Foundations of Social Research: Meaning and Perspective in the Research Process*. London: SAGE.
- Davis, C. (1995). Extensive reading: an expensive extravagance? *ELT Journal*, **49**(4): pp. 329-336.
- Day, R. (2003). *Extensive reading and the EFL curriculum* [Online]. Available from: <http://www.cape.edu/docs/TTalk0022.pdf> [Accessed: 09/07/ 2012].
- Day, R. and Bamford, J. (1998). *Extensive Reading in Second Language Classroom*. Cambridge: Cambridge University Press.
- Day, R. and Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, **14**(2): pp. 136-141.
- Day, R., Omura, C. and Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, **7**(2): pp. 541-551.
- Dekeyser, R. M. (2003). Implicit and explicit learning. In Long, M. H. and Doughty, C. J. (eds.) *The Handbook of Second Language Acquisition*. Oxford: Blackwell, pp. 313-348.
- Donnell, M. and Wood, M. (1999). *Becoming a Reader: A Developmental Approach to Reading Instructions*. Boston, MA: Allyn and Bacon.
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.
- Duffy, G. (1997). Powerful models or powerful teachers? An argument for teacher as entrepreneur. In Stahl, S. and Hayes, D. (eds.) *Instructional Models in Reading*. NJ: Lawrence Erlbaum Associates, Inc, pp. 351-366.
- Duffy, M. E. (1985). Designing nursing research: the qualitative-quantitative debate. *Journal of Advanced Nursing*, **10**(3): pp. 225-232.
- Dupuy, B. and Krashen, S. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning*, **4** (1): pp. 55-63.
- El-Dakhs, D. and Mitchell, A. (2012). *Spelling errors among EFL high-school graduates*. A paper presented at the 4th annual conference, Kingdom of Saudi Arabia Association of Language Teachers. Available from: http://www.ksaalt.org/conference_proceedings.html [Accessed: 06/01/2014].

- El-Hawat, A. (2003). Libyan education. In Teferra, D. and Altbach, P. G. (eds.) *African Higher Education: An International Reference Hand Book*. Bloomington: Indiana University Press, pp. 391-402.
- Elley, W. B. (1991). Acquiring Literacy in a second language: The effect of book-based programs. *Language learning*, **41**(3): pp. 375-411.
- Elley, W. B. and Mangubhai, F. (1981). *The Impact of a Book Flood in Fiji Primary Schools*. Wellington: New Zealand Council for Educational Research.
- Ellis, N. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review*, **11**: pp. 37-56.
- Ellis, R. (1999). *Learning a Second Language Through Interaction*. Amsterdam: John Benjamins Publishing Company.
- Ellis, R. and He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, **21**(2): pp. 285-301.
- Emhamed, E. and Krishnan, K. (2011). Investigating Libyan teachers' attitude towards integrating technology in teaching English in Sebha secondary schools. *Academic Research International*, **1**(3): pp. 182-192.
- Eskey, D. (1988). Holding in the bottom: An interactive approach to the language problems of second language readers. In Carrell, P., Devine, J. and Eskey, D. (eds.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press, pp. 93-100.
- Evered, E. (2012). *Empire and Education under the Ottomans: Politics, Reform and Resistance from the Tanzimat to the Young Turks*. London: IB Tauris.
- Faerch, C., Haastrup, K. and Phillipson, R. (1984). *Learner Language and Language Learning*. Clevedon: Multilingual Matters.
- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, **31**(2): pp. 105-119.
- Fareh, S. (2010). Challenges of teaching English in the Arab world: Why can't EFL programs deliver as expected? *Procedia-Social and Behavioral Sciences*, **2**(2): pp. 3600-3604.
- Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language*, **20**(1): pp. 19-42.

- Fraenkel, J. R. and Wallen, N. E. (2006). *How to Design and Evaluate Research in Education*. 6th. New York: McGraw-Hill.
- Freebody, P. and Anderson, R. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly*, **18**: pp. 277-294.
- Furuta, N. (2012). Learning L2 vocabulary through extensive reading: A case study. *Inquiry Report (ENG 732, Spring 2012)* [Online]. Available from: <http://naotoshifuruta.myefolio.com/Uploads/Extensive%20reading%20and%20L2%20vocabulary%20acquisition.pdf> [Accessed: 12/11/2013].
- García, G. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly*, **26**(4): pp. 371-392.
- Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, **9**(2): pp. 198-217.
- Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition*, **21**(2): pp. 319-333.
- Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, **6**(4): pp. 126-135.
- Goodman, K. S. (1976). Reading: A psycholinguistic guessing game. In Singer, H. and Ruddell, R. (eds.) *Theoretical Models and Processes of Reading*. 2nd ed. Newark, DE: International Reading Association, pp. 497-508.
- Goodman, K. S. (1988). The reading process. In Carrell, P., Devine, J. and Eskey, D. (eds.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press, pp. 11-21.
- Gough, P. B. (1972). One second of reading. In Kavanaugh, J. I. and Mattingly, I. G. (eds.) *Language by Eye and Ear*. Cambridge: MIT press, pp. 331-358.
- Goulden, R., Nation, P. and Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, **11**(4): pp. 341-363.
- GPCE (General People's Committee of Education). (2008). *The development of education: national report of Libya*. For the Session (48) of the International Conference on Education. Geneva, from 25-28 November 2008. Available from: http://www.ibe.unesco.org/National_Reports/ICE_2008/libya_NR08.pdf [Accessed: 06/02/2014].

- GPCE (General People's Committee of Education). (n. d). *The Libyan National Report on Adult learning and Literacy Education*. Available from: http://www.unesco.org/fileadmin/MULTIMEDIA/INSTITUTES/UIIL/confintea/pdf/National_Reports/Arab%20States/libya_en.pdf [Accessed: 16/3/2014].
- Grabe, W. (1988). Reassessing the term "interactive" In Carrell, P., Devine, J. and Eskey, D. (eds.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press, pp. 56-70.
- Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, **25**(3): pp. 375-406.
- Grabe, W. (2009). Reading in a second language. In Kaplan, R. B. (ed.) *The Oxford Handbook of Applied Linguistics*. New York: Oxford University Press, pp. 49-59.
- Grabe, W. and Stoller, F. (1997). Reading and vocabulary development in a second language: A case study. In Coady, J. and Huckin, T. (eds.) *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, pp. 98-122.
- Grabe, W. and Stoller, F. (2001). Reading for academic purpose: Guidelines for the ESL/EFL teacher. In Celce-Murcia, M. (ed.) *Teaching English as a Second or Foreign Language*. Boston: Heinle and Heinle, pp. 187-203.
- Grabe, W. and Stoller, F. (2002). *Teaching and Researching Reading*. Harlow: Longman.
- Grotjahn, R. (1987). On the methodological basis of introspective methods In Faerch, C. and Kasper, G. (eds.) *Introspection in Second Language Research*. Clevedon, UK: Multilingual Matters, pp. 54-81.
- Gulinao, A. (2013). *Libya educational system* [Online]. Available from: <http://www.slideshare.net/Rhaine/libyareportko> [Accessed: 12/02/2014].
- Haastrup, K. and Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics*, **10**(2): pp. 221-240.
- Hafiz, F. M. and Tudor, I. (1990). Graded readers as an input medium in L2 learning. *System*, **18**(1): pp. 31-42.
- Hamdy, A. (2007). *ICT in education in Libya*. Libya Country Report. Available from: <http://www.infodev.org/en/Document.412.pdf> [Accessed: 23/12/2010].
- Hargis, C. (1999). *Teaching and Testing in Reading: A Practical Guide for Teachers And Parents*. Illinois: Charles Thomas Publishers, Inc.

- Hartas, D. (2010). Educational research and inquiry: Key issues and debates. In Hartas, D. (ed.) *Educational Research and Inquiry: Quantitative And Qualitative Approaches*. London: Continuum, pp. 13-32.
- Hedgcock, J. S. and Ferris, D. R. (2009). *Teaching Readers of English: Students, Texts, and Contexts*. New York: Routledge.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, **21**(2): pp. 303-317.
- Hill, D. R. (1997). Setting up an extensive reading programme: Practical tips. *Language Teacher-Kyoto-JALT*, **21**: pp. 17-20.
- Hill, D. R. (2008). Graded readers in English. *ELT Journal*, **62**(2): pp. 184-204.
- Honeycutt, J. M. and Bryan, S. P. (2010). *Scripts and Communication for Relationships*. New York: Peter Lang Publishin.
- Hong, X. (2010). Review of effects of glosses on incidental vocabulary learning and reading comprehension. *Chinese Journal of Applied Linguistics* **33**(1): pp. 56-73.
- Hopkins, G. (1997). Sustained silent reading helps develop independent readers (and writers). Available from: http://www.educationworld.com/a_curr/curr038.shtml [Accessed: 03/10/2010].
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, **61**(3): pp. 355-382.
- Horst, M., Cobb, T. and Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, **11**(2): pp. 207-223.
- Horst, M. and Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Review*, **56**(2): pp. 308-328.
- Howatt, A. P. (1984). *A History of English Language Teaching*. Oxford: Oxford University Press.
- Hu, M. and Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, **13**(1): pp. 403-430.
- Huberman, A. M. and Miles, M., B (1998). Data Management and analysis methods. In Denzin, N. K. and Lincoln , Y. S. (eds.) *Collecting and Interpreting Qualitative Materials*. California: SAGE, pp. 179-210.

- Huckin, T. and Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, **21**(02): pp. 181-193.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental learning. In Arnaud, P. J. and Bejoint, H. (eds.) *Vocabulary in Applied Linguistics*. London: Macmillan, pp. 113-125.
- Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal*, **77**(2): pp. 139-147.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal, rehearsal, and automaticity. In Robinson, P. (ed.) *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, pp. 258-286.
- Hulstijn, J. H. (2003). Incidental and Intentional Learning. In Doughty, J. C. and Long, M. H. (eds.) *The Handbook of Second Language Acquisition*. London: Blackwell Publishing Ltd, pp. 349-381.
- Hulstijn, J. H., Hollander, M. and Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, **80**(3): pp. 327-339.
- Hulstijn, J. H. and Laufer, B. (2001). Some empirical evidence for the involvement Load Hypothesis in vocabulary acquisition. *Language Learning*, **51**(3): pp. 539-558.
- Hunt, A. and Beglar, D. (1998). Current research and practice in teaching vocabulary. *The Language Teacher Online* [Online], **22**. Available from: http://jalt-publications.org/old_tlt/articles/1998/01/hunt [Accessed: 14/2/2013].
- Ivankova, N. and Creswell, J. (2009). Mixed methods. In Heigham, J. and Croker, R. (eds.) *Qualitative Research in Applied Linguistics: A Practical Introduction*. New York: Palgrave Macmillan, pp. 135-161.
- Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language*, **20**(1): pp. 70-91.
- Jacoby, L., Craik, F. and Begg, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, **18**(5): pp. 585-600.
- Janopoulos, M. (1986). The relationship of pleasure reading and second language writing proficiency. *TESOL Quarterly*, **20**(4): pp. 763-768.

- Jdetawy, L. (2011). Problems encountered by Arab EFL learners. *Language in India*, **11**(3): pp. 19-27.
- Jenkins, J. R. and Dixon, R. (1983). Vocabulary learning. *Contemporary Educational Psychology*, **8**(3): pp. 237-260.
- Jenkins, J. R., Pany, D. Z. and Schreck, J. V. (1978). *Vocabulary and reading comprehension: Instructional effects. (Technical Report No. 100)*. University of Illinois, Center for the Study of Reading.
- Jenkins, J. R., Stein, M. L. and Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, **21**(4): pp. 767-787.
- Juel, C. (1999). The messenger may be wrong, but the Message may be Right. In Oakhill, J. and Beard, S. (eds.) *Reading Development and the Teaching of Reading*. Malden, MA: Blackwell, pp. 201-212.
- Kaivanpanah, S. and Alavi, M. (2008). Deriving unknown word meaning from context is it reliable? *RELC Journal*, **39**(1): pp. 77-95.
- Kerka, S. (2000). *Incidental learning: Trends and Issues*. Alert 18. Educational Resources Information. Available from: <http://files.eric.ed.gov/fulltext/ED446234.pdf> [Accessed: 04/07/2012].
- Khan, I. (2012). Teaching and learning of English: An exploratory study of difficulties and strategies. *British Journal of Social Sciences*, **1**(1): pp. 76-93.
- Kim, Y. H. and Goetz, E. T. (1994). Context effects on word recognition and reading comprehension of poor and good readers: A test of the interactive-compensatory hypothesis. *Reading Research Quarterly*, **29**(1): pp. 179-188.
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, **78**(3): pp. 285-299.
- Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. New York: Longman.
- Krashen, S. (1989a). *Language Acquisition and Language Education*. London: Prentice Hall International.
- Krashen, S. (1989b). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, **73**(4): pp. 440-464.

- Krashen, S. (1991). The input hypothesis: An update. In James, E. (ed.) *Georgetown University Round Table on Languages and Linguistics*. Washington, D.C: Georgetown University Press, pp. 409-431.
- Krashen, S. (1993). *The Power of Reading: Insights from the Research*. Colorado: Eaglewood Libraries Unlimited.
- Krashen, S. (1997). The comprehension hypothesis: Recent evidence. *English Teachers' Journal* **51**: pp. 17-29.
- Krashen, S. (2004). *Applying the comprehension hypothesis: Some suggestions*. Paper presented at the 13th International Symposium and Book Fair on Language Teachingair, English Teachers Association of the Republic of China. Taipei, Taiwan. Available from: http://www.sdkrashen.com/articles/eta_paper/all.html [Accessed: 19/02/2014].
- Kusanagi, Y. (2004). The class report 2: Course evaluation of pleasure reading course. *The Journal of Rikkyo University Language Center*, **11**: pp. 29-42.
- Laberge, D. and Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, **6**(2): pp. 293-323.
- Lai, F. K. (1993). Effect of extensive reading on English learning in Hong Kong. *CUHK Education Journal*, **21**: pp. 23-36.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In Lauren, C. and Nordman, L. (eds.) *Special Language: From Human Thinking to Thinking Machines*. Clevedon: Multilingual Matters, pp. 316-323.
- Laufer, B. (1992). How much Lexis is necessary for reading comprehension? In Arnauld, P. and Bejoint, H. (eds.) *Vocabulary and Applied Linguistics*. London: Macmillan, pp. 126-132.
- Laufer, B. (1997). The lexical plight in second language reading. In Coady, J. and Huckin, T. (eds.) *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge: Cambridge University Press, pp. 20-34.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, **19**(2): pp. 255-271.
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, **59**(4): pp. 567-587.

- Laufer, B. and Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, **54**(3): pp. 399-436.
- Laufer, B. and Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, **22**(1): pp. 1-26.
- Laufer, B. and Nation, I. S. P. (2012). Vocabulary. In Gass, S. and Mackey, A. (eds.) *The Routledge Handbook of Second Language Acquisition*. New York: Routledge, pp. 163-176.
- Laufer, B. and Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, **48**(3): pp. 365-391.
- Laufer, B. and Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research*, **15**(4): pp. 391-411.
- Lewis, M. (1993). *The Lexical Approach: The State of ELT and a Way Forward*. Hove, England: Language Teaching Publications.
- LMHE (The Libyan Minister of Higher Education). (2010). *Education at the university level*. Available from: http://eacea.ec.europa.eu/tempus/participating_countries/overview/libya/overview_of_hes_final.pdf [Accessed: 19/01/2014].
- LNTO (Libya National Tempus Office). (2012). *Higher Education in Libya*. Available from: file:///E:/libya_hier%20education%20overview_of_hes_final.pdf [Accessed: 21/01/2014].
- Macalister, J. (2010). Investigating teacher attitudes to extensive reading practices in higher education: Why isn't everyone doing it? *RELC Journal*, **41**(1): pp. 59-75.
- Mackey, A. and Gass, S. M. (2005). *Second Language Research: Methodology and Design*. Mahwah: Lawrence Erlbaum Associates.
- Maley, A. (1999). Surviving the 20th century. *English Teaching Professional*, **10**: pp. 3-7.
- Martinez-Lage, A. (Year) Published. Benefits of keeping a reading journal in the development of second language reading ability. In Terry, R., (ed). *The Future is Now*. Dimension: Language '95, 1995 Roswell, GA. Southern Conference on Language Teaching, 65-79.

- Mason, B. and Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, **25**(1): pp. 91-102.
- Masuhara, H., Kimura, T., Fukada, A. and Takeuchi, M. (1996). Strategy training or/and extensive reading? In Hickey, T. and Williams, J. (eds.) *Language, Education, and Society in a Changing World*. Clevedon, UK: Multilingual Matters, pp. 263-274.
- Mclaughlin, B., Rossman, T. and Mcleod, B. (1983). Second language learning: An information processive perspective. *Language Learning*, **33**(2): pp. 135-158.
- Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics: Abstracts*, **13**(4): pp. 221-246.
- Meara, P. (1996). *The vocabulary knowledge framework* [Online]. Available from: <http://www.lognostics.co.uk/vlibrary/meara1996c.pdf> [Accessed: 17/04/2012].
- Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In Schmitt, N. and McCarthy, M. (eds.) *Vocabulary: Discription, Acquisition, and Pedagogy*. Campridge: Campridge University Press, pp. 84-102.
- Melka Teichroew, F. J. (1982). Receptive versus productive vocabulary: A survey. *Interlanguage Studies Bulletin*, **6**: pp. 5-33.
- Mikulecky, B. (1990). *A Short Course in Teaching Reading Skills*. New York: Addison-Wesley Publishing Company. Inc.
- Milton, J. (2008). Vocabulary uptake from informal learning tasks. *Language Learning Journal*, **36**(2): pp. 227-237.
- Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Mondria, J. A. and Wit-De Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics*, **12**(3): pp. 47-57.
- Montgomery, D. C. (2005). *Design and Analysis of Experiments*. New York: Wiley.
- Mourtaga, K. R. (2006). Some reading problems of Arab EFL students. *Journal of Al-Aqsa University*, **10**(2): pp. 75-91.

- Nagy, W. (1997). On the role of context in first-and second-language vocabulary learning. In Schmitt, N. and McCarthy, M. (eds.) *Vocabulary Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 64-83.
- Nagy, W. and Anderson, R. (1984). How many words are there in printed school English? *Reading Research Quarterly*, **19**: pp. 304-330.
- Nagy, W., Anderson, R. and Herman, P. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, **24**(2): pp. 237-270.
- Nagy, W. and Herman, P. (1985). Incidental vs. instructional approaches to increasing reading vocabulary. *Educational Perspectives*, **23**(1): pp. 16-21.
- Nagy, W. and Herman, P. (1987). Breadth and depth of vocabulary knowledge: Implications of acquisition and instruction. In Mckeown, M. G. and Curtis, M. E. (eds.) *The Nature of Vocabulary Acquisition*. Hillsdale, NJ: Erlbaum, pp. 19-36.
- Nagy, W., Herman, P. and Anderson, R. (1985). Learning words from context. *Reading Research Quarterly*, **20**(2): pp. 233-253.
- Najeeb, S. (2012). The business of teaching English as a second language: A Libyan case study. *Procedia-Social and Behavioral Sciences*, **70**: pp. 1243-1253.
- Najeeb, S. and Eldokali, E. (n. d). *English in Libya: The language of development* [Online]. Available from: http://www.iairs.org/Abstracts_english/INV1.pdf [Accessed: 18/10/2013].
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.
- Nation, I. S. P. (1997). The language learning benefits of extensive reading. *The Language Teacher*, **21**(5): pp. 13-16.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2005a). Reading faster. *PSSSA*, **36**: pp. 21-37.
- Nation, I. S. P. (2005b). Vocabulary learning through extensive reading. In Poedjosoedarmo, G. (ed.) *Innovative Approaches to Reading and Writing*. Singapore: RELC Anthology Series, pp. 10-21.

- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, **63**(1): pp. 59-82.
- Nation, I. S. P. (2010). *Learning Vocabulary in Another Language*. 2nd ed. Cambridge: Cambridge University Press.
- Nation, I. S. P. and Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, **12**(2): pp. 355-380.
- Nation, I. S. P. and Waring, R. (1997). Vocabulary size, text coverage and word lists. In Schmitt, N. and McCarthy, M. (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 6-19.
- Nishino, T. (2007). Beginning to read extensively: A case study with Mako and Fumi. *Reading in a Foreign Language*, **19**(2): pp. 76-105.
- Nunan, D. (1992). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.
- Nunan, D. (1998). Teaching grammar in context. *ELT Journal* **52**(2): pp. 101-109.
- Nunan, D. and Bailey, K. M. (2009). *Exploring Second Language Classroom Research: A Comprehensive Guide*. Boston: Heinle Cengage Learning.
- Nuttall, C. (1996). *Teaching Reading Skills in a Foreign Language*. Oxford: Heinemann.
- O'sullivan, A. (2004). *Reading and Arab college students: Issues in the United Arab Emirates Higher Colleges of Technology*. In Proceedings of the First International Online Conference on Second and Foreign Language Teaching and Research: Beyond Borders. Symposium conducted at First International Online Conference on Second and Foreign Language Teaching and Research. Available from: <http://www.readingmatrix.com/conference/pp/proceedings/sullivan.pdf> [Accessed: 13/06/2012]
- Oakhill, J. and Beard, R. (1999). *Reading Development and the Teaching of Reading: A Psychological Perspective*. Oxford: Blackwell.
- Obeidi, A. (2001). *Political Culture in Libya*. Surrey: Routledge.
- Opie, C. (2004). *Doing Educational Research*. London: SAGE.
- Orafi, S. M. S. and Borg, S. (2009). Intentions and realities in implementing communicative curriculum reform. *System*, **37**(2): pp. 243-253.

- Oxford, R. and Scarcella, R. (1994). Second language vocabulary learning among adults: State of the art in vocabulary instruction. *System*, **22**(2): pp. 231-243.
- Paran, A. (1996). Reading in EFL: Facts and fictions. *ELT Journal*, **50**(1): pp. 25-34.
- Paribakht, T. S. and Wesche, M. B. (1999). Reading and “incidental” L2 vocabulary acquisition. *Studies in Second Language Acquisition*, **21**(02): pp. 195-224.
- Parry, K. (1991). Building a vocabulary through academic reading. *Tesol Quarterly*, **25**(4): pp. 629-653.
- Pathan, M. and Al-Dersi, Z. (2013). Investigating the role of short-stories in overcoming the problems faced by the Libyan EFL learners in reading comprehension skill. *The Criterion Journal*(12): pp. 1-8.
- Patil, Z. N. (2010). Arab learners of English and reverse visualization as a reading problem. *Arab World English Journal*, **1**(3): pp. 1-23.
- Pellicer-Sanchez, A. and Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do "Things Fall Apart"? *Reading in a Foreign Language*, **22**(1): pp. 31-55.
- Pigada, M. and Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, **18**(1): pp. 1-28.
- Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers*, **4**: pp. 37-45.
- Pimsleur, P. (1967). A memory schedule. *The Modern Language Journal*, **51**(2): pp. 73-75.
- Pitts, M., White, H. and 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): pp. 271-275.
- Powell, S. (2005). Extensive reading and its role in Japanese high schools. *The Reading Matrix*, **5**(2): pp. 28- 42.
- Prowse, P. (2002). Top ten principles for teaching extensive reading--a response. *Reading in a Foreign Language*, **14**(2): pp. 142-145.
- Pulido, D. (2004). The effect of cultural familiarity on incidental vocabulary acquisition through reading. *The Reading Matrix*, **4**(2): pp. 20–53.

- Punch, M. (1994). Politics and ethics in qualitative research. In Denzin, N. K. and Lincoln, Y. S. (eds.) *Handbook of Qualitative Research*. Thousand Oaks: SAGE, pp. 83-97.
- Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, **52**(3): pp. 513-536.
- Rajendran, S. (2010). ELT in Libyan universities: A pragmatic approach. *Language in India*, **10**(9): pp. 63-71.
- Raptis, H. (1997). Is second language reading vocabulary best learned by reading? *Canadian Modern Language Review*, **53**(3): pp. 566-580.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, **62**(8): pp. 1457-1506.
- Rayner, K. and Pollatsek, A. (1989). *The Psychology of Reading*. Englewood Cliffs NJ: Prentice Hall.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Reed, S. K. (2007). *Cognition: Theory and Applications*. USA: Thomson Wadsworth.
- Remenyi, D., Williams, B., Money, A. and Swartz, E. (1998). *Doing Research in Business and Management. An Introduction to Process and Method*. London: SAGE.
- Richards, J. (1976). The role of vocabulary teaching. *TESOL Quarterly*, **10**(1): pp. 77-89.
- Richards, J. and Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. Malaysia: Pearson Education.
- Rieder, A. (2003). Implicit and explicit learning in incidental vocabulary acquisition. *VIEWS, Vienna English Working Papers*, **12**(2): pp. 24-39.
- Robb, T. and Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, **5**(2): pp. 239-251.
- Rott, S. (1999a). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, **21**: pp. 589-619.
- Rott, S. (1999b). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, **21**: pp. 289-619.

- Rumelhart, D. (1977). Toward an interactive model of reading. In Dornic, S. (ed.) *Attention and Performance IV*. New York: Academic Press, pp. 573-603.
- Rumelhart, D. (1984). Understanding reading In Flood, J. (ed.) *Understanding Reading Comprehension*. New York: International Reading Association, pp. 1-20.
- Rumelhart, D. (1990). Toward an interactive model of reading. In Dornic, S. (ed.) *Attention and Performance*. New York: Academic Press, pp. 573-603.
- Ryan, A. and Meara, P. (1991). The case of the invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language*, 7(2): pp. 531-531.
- Saaïd, N. (2010). *The use of technology (audiovisual aids) in teaching English as a foreign language in Benghazi secondary schools*. MA Dissertation, Leicester University.
- Samuels, S. J. and Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 13(2): pp. 107-121.
- Samuels, S. J. and Kamil, M. L. (1988). Models of reading process. In Carrell, P., Devine, J. and Eskey, D. (eds.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press, pp. 22-36.
- Saragi, T., Nation, P. and Meister, G. (1978). Vocabulary Learning and Reading. *System*, 6(2): pp. 72-78.
- Saunders, M., Lewis, P. and Thornhill, A. (2009). *Research Methods for Business Students*. 5th ed. Harlow: Pearson Education.
- Saunders, M., Lewis, P. and Thornhill, A. (2012). *Research Methods for Business Students*. 6th ed. Harlow: Pearson Education.
- Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly*, 18(2): pp. 199-219.
- Sawani, F. (2009). *Factors affecting English teaching and its materials preparation in Libya*. PhD Thesis, University of Essex.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2): pp. 129-158.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11: pp. 11-26.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

- Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N. and McCarthy, M. (eds.) (1997). *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press.
- Schmitt, N. and Meara, P. (1997). Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition*, **19**(1): pp. 17-36.
- Schmitt, N., Schmitt, D. and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, **18**(1): pp. 55-88.
- Schmitt, N. and Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, **36**(2): pp. 145-171.
- Schouten-Van Parreren, C. (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts? In Nation, P. and Carter, R. (eds.) *Vocabulary Acquisition*. Amsterdam: : Free University Press (AILA Review/Revue de l'AILA 6), pp. 75-85.
- Shefelbine, J. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior*, **22**(1): pp. 71-97.
- Shihiba, S. (2011). *An investigation of Libyan EFL teachers' conceptions of the communicative learner-centred approach in relation to their implementation of an English language curriculum innovation in secondary schools*. PhD Thesis, Durham University.
- Shuell, T. J. (1996). Teaching and learning in a classroom context. In Berliner, D. C. and Calfee, R. (eds.) *Handbook of Educational Psychology*. New York: Macmillan, pp. 726-764.
- Smith, F. (1971). *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read*. New York: Holt, Rinehart and Winston.
- Sökman, A. (1997). Current trends in teaching second language vocabulary'. In Schmitt, N. and McCarthy, M. (eds.) *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press, pp. 237-257.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, **16**(1): pp. 32-71.

- Taghavi, M. (2013). *A critical analysis of Higher Education sector in Libya: A socio-economic analysis* [Online]. Available from: http://works.bepress.com/cgi/viewcontent.cgi?article=1045&context=mr_farzanegan [Accessed: 17/01/2014].
- Taguchi, E., Takayasu-Maass, M. and Gorsuch, G. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language*, **16**(2): pp. 70-96.
- Tamtam, A., Gallagher, F., Olabi, A. and Naher, S. (2011). Higher education in Libya, system under stress. *Procedia-Social and Behavioral Sciences*, **29**: pp. 742-751.
- Tantani, A. S. (2012). *Significant relationships between EFL teachers' practice and knowledge in the teaching of grammar in Libyan secondary schools*. PhD Thesis, University of Sunderland.
- Thomas, R. M. (1998). *Conducting Educational Research: A comparative view*. Westport: Bergin & Garvey.
- Thompson-Panos, K. and Thomas-Ruzic, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quarterly*, **17**(4): pp. 609-623.
- Tracey, D. H. and Morrow, L. M. (2006). *Lenses on Reading : An Introduction to Theories and Models*. New York: Guilford Press.
- Tsang, W. K. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics*, **17**(2): pp. 210-233.
- UNESCO (1968). *English teaching: Libya, (mission) 27 August 1965 - 30 June 1968*. Available from: <http://unesdoc.unesco.org/images/0000/000087/008728eb.pdf> [Accessed: 21/10/ 2013].
- Urquhart, S. and Weir, C. (1998). *Reading in a Second Language: Process, Product, and Practice*. London: Longman.
- Wang, A., Thomas, M., Inzana, C. and Primicerio, L. (1993). Long-term retention under conditions of intentional learning and the keyword mnemonic. *Bulletin of the Psychonomic Society*, **31**: pp. 545-545.
- Waring, R. (2011). Extensive reading in English teaching. In Widodo, H. and Cirocki, A. (eds.) *Innovation and Creativity in ELT methodology*. New York: Nova Publishers, pp. 69-80
- Waring, R. and Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English Speaking World*, **4**: pp. 11-23.

- Waring, R. and 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): pp. 130-163.
- Webb, S. (2005). Receptive and productive vocabulary learning. *Studies in Second Language Acquisition*, **27**: pp. 33-52.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, **28**(1): pp. 46–65.
- Webb, S. (2008a). The effect of context on incidental vocabulary learning. *Reading in a Foreign Language*, **20**(2): pp. 323-245.
- Webb, S. (2008b). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, **30**(1): pp. 79-95.
- Wellington, J. (1996). *Methods and Issues in Educational Research*. Sheffield: University of Sheffield Division of Education.
- Wellington, J. (2000). *Educational Research: Contemporary Issues and Practical Approaches. 4th*. London: Continuum.
- Wesche, M. and Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, **53**(1): pp. 13-40.
- Wesche, M. and Paribakht, T. S. (1999). Introduction. *Studies in Second Language Acquisition*, **21**: pp. 175-180.
- White, T. G., Graves, M. F. and Slater, W. H. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. *Journal of Educational Psychology*, **82**(2): pp. 281-290.
- Wilkinson, D. (2010). *Extensive reading: What is it and why should we be doing it?* [Online]. ESL/EFL/TEFL Teaching Information. Available from: http://www.totalesl.com/e_articles_print.php?id=536 [Accessed: 09/07 2011].
- Wode, H. (1999). Incidental vocabulary acquisition in the foreign language classroom. *Studies in Second Language Acquisition*, **21**(2): pp. 243-258.
- Yamamoto, Y. (2011). Bridging the gap between receptive and productive vocabulary size through extensive reading. *The Reading Matrix*, **11**(3): pp. 226-242.
- Yamashita, J. and Ichikawa, S. (2010). Examining reading fluency in a foreign language: Effects of text segmentation on L2 readers. *Reading in a Foreign Language*, **22**(2): pp. 263-283.

- Yang, L. (1997). Tracking the acquisition of L2 vocabulary: The Keki Language experiment. In Coady, J. and Huckin, T. (eds.) *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, pp. 125- 156.
- Yin, R. K. (1994). *Case Study Research: Design and Methods*. 2nd ed. Thousand Oaks, California: SAGE.
- Yin, R. K. (2009). *Case Study Research: Design and Methods*. 4th ed. Thousand Oaks, California: SAGE.
- Zainal, Z. (2003). Critical review of reading model and theories in first and second language. *Jurnal Kemanusiaan* [Online]. Available from: http://eprints.utm.my/758/1/ZaidahZainal2003_CriticalReviewofReadingModel.pdf [Accessed: 28/9/2013].

Appendices

Appendix 1: Oxford bookworm levels tests (stages 3, 4, and 5)

Appendix 1 (A). Oxford bookworm stage 3 level test

Stage 3

Student name: _____

We had a small party afterwards, just the four of us. Then Ray and Phil went home and Jenny and I were alone together.

'Jenny, we're really married!'

'Yes. Now I can be as terrible to you as I like!'

For three years we had to make every dollar do the work of two. All through the summer holidays we worked at the Boat Club in Dennis Port. It _____ (is, was, were, will be) hard work, but we were never _____ (too, to, so, very) tired to be kind to each _____ (another, us, one, other.) I say 'kind' because there are _____ (none, know, no, nothing) words to describe our love and _____ (happy, happiness, happily, happier) together.

After the summer we found _____ (a, the, an, that) 'cheap' flat near the university. It _____ (has been, had been, was, could be) on the top floor of an _____ (new, old, modern) house and was actually very expensive. _____ (So, Although, But, Because) what could we do? There weren't _____ (much, many, some, lots) flats around.

'Hey, preppie,' said Jenny _____ (when, while, during, at) we arrived there. 'Are you my _____ (brother, husband, wife, father) or aren't you?'

'Of course I'm _____ (you, you're, yours, your) husband.'

'Show me, then.' (My god, _____ (I, I'm, I was, I'll) thought, in the street?)

'Carry me _____ (into, in, at, onto) our first home!'

I carried her _____ (over, high, up, above) the five steps to the front

_____ (window, cupboard, door, desk)

'Why did you stop?' she asked. _____ (This, These, It, The) isn't our home.

Upstairs Preppie!

There _____ (were, was, had, did) twenty-four stairs up to our

_____ (room, house, office, flat) and I had to stop half-way.

_____ (What, When, Why, Where) are you so heavy?' I asked

_____ (she, her, him, he,) 'Perhaps I'm expecting a baby.'

My _____ (Dog!, God!, Bag!, Dad!) Are you?'

'Ha! I frightened you _____ (just, when, then, soon) didn't I?'

'Well, yes, just for _____ (a, an, the, one) second or two.'

I carried her _____ (a, to, the, some) rest of the way. There were _____ (some,

very, not, many) few moments in those days when _____ (she, us, we, our)

were not worrying about money. Very _____ (little, small, many, few) and very

wonderful – and that moment _____ (was, has been, is, will be) one of them.

Appendix 1 (B). Oxford bookworm stage 4 level test

Stage 4

Student name _____

It all started when the big, white bird flew out of the shiny leaves and yellow flowers. It rose up suddenly and turned away towards the mountains. I followed it. What else could I do in the middle of such a bright April day, at the foot of the White Mountains of Crete? The road was hot and dusty, but the valley was _____ (blue, green, red, yellow) and full of the sound of _____ (trees., water., rocks., flowers.) The white wings which flew before _____ (I, my, me, mine) moved quickly in and out of _____ (the, that, this, these) deep shadow of the trees and _____ (a, an, some, the) air was full of the sweet _____ (taste, sound, feel, smell) of the lemon grove.

The car _____ (of, past, from, about) Heraklion had stopped where the path _____ (for, of, at, in) Agios Georgios leaves the road. I _____ (get, got, gotten, getting) out and turned to thank the _____ (USA, America, American, Americans) couple who had brought me this _____ (long., far., much., time.) Mrs Studebaker looked out of the _____ (car, train, van, bus) window. 'But are you going to _____ (being, been, to be, be) all right? You're sure this is _____ (a, the, to, some) right place? What does that sign _____ (tell?, talk?, ask?, say?)

The sign was in Greek. 'It's _____ (all, OK, everything, not) right,' I laughed. 'That's 'Agios Georgios', _____ (then, also, so that, and) the village is not far away, _____ (downwards, down, bellow, under) this path.'

I had been in _____ (Athens, Paris, New York, London) since January of the year before. _____ (I'm, I've, I, I'll) worked as a very unimportant secretary _____ (at, to, of, on) the British Embassy. I had always _____ (want, wanted, wanting, to want) to visit Greece, and thought I _____ (were, am, was being, was) lucky, at the age of _____ (one-twenty, twenty-one, twentyone, twenty first)) to get any kind of job _____ (there., where., somewhere., anywhere.) I had enjoyed my time in _____ (Greece, Athens, Heraklion, Agios Georgios) and had worked hard to learn _____ (Greece., Greeks., language., Greek.)

I was going to spend Easter _____ (with, at, to, together) my cousin, Frances Scorby. She was _____ (come, came, coming, had come) with some friends, who had hired _____ (a, an, the, one) boat, but she was going to leave them for a few days to be in Crete with me. We would join her friends later.

Appendix 1 (C). Oxford bookworm stage 5 level test

Stage 5

Student name:-----

That is how Moore arrived at Benchurch. It was a market town, and once a week it was quite busy for a few hours. The rest of the time it was a very quiet, sleepy little place. Moore spent his first night at the only hotel in the town. The landlady was very kind and helpful, but the hotel was not really quiet enough for him. The second day he started looking for a house to rent.

There was _____ (alone, solo, only, solitary) one place that he liked. It _____ (was, was being, had been, is) more than quiet – it was deserted _____ (and, also, as well, but) very lonely. It was a big, _____ (new, modern, elderly, old) seventeenth-century house. It had tiny windows _____ (like, similar, the same, alike) a prison, and a high brick _____ (fence, hedge, wall, building) all round it. It would be _____ (firm, hard, stiff, solid) to imagine a more unwelcoming place. _____ (Although, But, While, Despite) it suited Moore perfectly. He went _____ (to, for, too, of) find the local lawyer, who was _____ (in charge, reliable, to blame, responsible) for the house.

Mr Camford, the _____ (solicitor, landlord, owner, lawyer) was very happy to rent the _____ (flat, house, cottage, castle) to him.

'I'd be glad to _____ (let, allow, enable, permit) you have it free,' he said, _____ (hardly, just, really, exactly) to have somebody living in it _____ (over, repeat, again, some more) after all these years. It's been _____ (empty, bare, vacant, clear) so long that people have spread _____ (an, some, the, a) lot of foolish stories about it. _____ (You can, You're going to, You're, You'll) be able to prove that the _____ (stories are, stories, story's, story) are wrong.'

Moore did not think _____ (it, that, of, about) was necessary to ask the lawyer _____ (of, to, for, with) more details of the 'foolish stories'. _____ (He's, He, He'll, He'd) paid his rent, and Mr Camford _____ (took, said, gave, asked) him the name of an old _____ (slave, servant, worker, maid) to look after him. He came _____ (by, up, over, away) from the lawyer's office with the _____ (locks, keys, codes, number) of the house in his pocket. _____ (He, She, They, It) then went to Mrs Wood, the _____ (director, boss, landlord, landlady) of the hotel.

'I'm renting a _____ (apartment, accommodation, building, house) for a few weeks,' he said. _____ (Shall, Must, Should, Can) you advise me about shopping, please? What do you think I shall need?'

'Where are you going to stay, sir?' the landlady asked. Moore told her.

She threw up her hands in horror. 'Not the Judge's House!' she said, and she grew pale as she spoke.

He asked her to tell him more about the house. 'Why is it called the Judge's House?' he said, 'and why doesn't anyone want to live in it?'

Appendix 2: **Instant book report**

Student's ID No: _____

Book title _____

We would like to ask you some questions about your reading of this story. Please be honest and answer the questions below. Data from this survey will be used for research purposes only and will not affect your university grades. Please put a tick (✓) in the box to show your answers and use the provided spaces for the questions which require complete answers.

Please feel free to use “OTHER” space to write any other issues or additional information of the following questions.

- Did you read the whole book? yes No

If not, how many pages have you read? _____

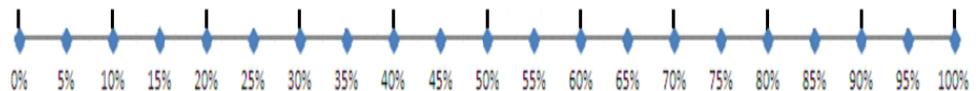
- Did you enjoy the book? yes No

If yes, why did you enjoy it?

If not, why didn't you enjoy it?

- Was it the first time you have read this story, or had you read it before either in English or any other language including Arabic?
 yes No

- Please rate your understanding of the story



Appendix 3: Schmitt's (2000) Vocabulary Levels Test

Vocabulary Levels Test: Version 1

This is one of two versions of the Vocabulary Levels Test that are available. See Schmitt, Schmitt, & Clapham for details on the procedures used to validate this test, and Schmitt and Schmitt (under review) for Version 2.

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning. Here is an example.

- 1. business
- 2. clock _____ part of a house
- 3. horse _____ animal with four legs
- 4. pencil _____ something used for writing
- 5. shoe
- 6. wall

You answer it in the following way.

- 1. business
- 2. clock 6 part of a house
- 3. horse 3 animal with four legs
- 4. pencil 4 something used for writing
- 5. shoe
- 6. wall

Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, these words are *business*, *clock*, and *shoe*.

If you have no idea about the meaning of a word, do not guess. But if you think you might know the meaning, then you should try to find the answer.

Version 1: The 2,000-word level

- 1. birth
- 2. dust _____ game
- 3. operation _____ winning
- 4. row _____ being born
- 5. sport
- 6. victory

- 1. choice
- 2. crop _____ heat
- 3. flesh _____ meat
- 4. salary _____ money paid regularly for doing a job
- 5. secret
- 6. temperature

1. cap
2. education _____ teaching and learning
3. journey _____ numbers to measure with
4. parent _____ going to a far place
5. scale
6. trick

1. attack
2. charm _____ gold and silver
3. lack _____ pleasing quality
4. pen _____ not having something
5. shadow
6. treasure

1. cream
2. factory _____ part of milk
3. nail _____ a lot of money
4. pupil _____ person who is studying
5. sacrifice
6. wealth

1. adopt
2. climb _____ go up
3. examine _____ look at closely
4. pour _____ be on every side
5. satisfy
6. surround

1. bake
2. connect _____ join together
3. inquire _____ walk without purpose
4. limit _____ keep within a certain size
5. recognize
6. wander

1. burst
2. concern _____ break open
3. deliver _____ make better
4. fold _____ take something to someone
5. improve
6. urge

- | | |
|-------------|--------------------------|
| 1. original | |
| 2. private | _____ first |
| 3. royal | _____ not public |
| 4. slow | _____ all added together |
| 5. sorry | |
| 6. total | |

- | | |
|-------------|----------------------|
| 1. brave | |
| 2. electric | _____ commonly done |
| 3. firm | _____ wanting food |
| 4. hungry | _____ having no fear |
| 5. local | |
| 6. usual | |

Version 1: The 3,000-word level

- | | |
|--------------|--|
| 1. belt | |
| 2. climate | _____ idea |
| 3. executive | _____ inner surface of your hand |
| 4. notion | _____ strip of leather worn around the waist |
| 5. palm | |
| 6. victim | |

- | | |
|--------------|---------------------------------|
| 1. acid | |
| 2. bishop | _____ cold feeling |
| 3. chill | _____ farm animal |
| 4. ox | _____ organization or framework |
| 5. ridge | |
| 6. structure | |

- | | |
|-------------|-------------------------|
| 1. bench | |
| 2. charity | _____ long seat |
| 3. jar | _____ help to the poor |
| 4. mate | _____ part of a country |
| 5. mirror | |
| 6. province | |

- | | |
|---------------|--------------------------------------|
| 1. boot | |
| 2. device | _____ army officer |
| 3. lieutenant | _____ a kind of stone |
| 4. marble | _____ tube through which blood flows |
| 5. phrase | |
| 6. vein | |

1. apartment
2. candle _____ a place to live
3. draft _____ chance of something happening
4. horror _____ first rough form of something written
5. prospect
6. timber

1. betray
2. dispose _____ frighten
3. embrace _____ say publicly
4. injure _____ hurt seriously
5. proclaim
6. scare

1. encounter
2. illustrate _____ meet
3. inspire _____ beg for help
4. plead _____ close completely
5. seal
6. shift

1. assist
2. bother _____ help
3. condemn _____ cut neatly
4. erect _____ spin around quickly
5. trim
6. whirl

1. annual
2. concealed _____ wild
3. definite _____ clear and certain
4. mental _____ happening once a year
5. previous
6. savage

1. dim
2. junior _____ strange
3. magnificent _____ wonderful
4. maternal _____ not clearly lit
5. odd
6. weary

Version 1: The 5,000-word level

1. balloon
2. federation _____ bucket
3. novelty _____ unusual interesting thing
4. pail _____ rubber bag that is filled with air
5. veteran
6. ward

1. alcohol
2. apron _____ stage of development
3. hip _____ state of untidiness or dirtiness
4. lure _____ cloth worn in front to protect your clothes
5. mess
6. phase

1. apparatus
2. compliment _____ expression of admiration
3. ledge _____ set of instruments or machinery
4. revenue _____ money received by the government
5. scrap
6. tile

1. bulb
2. document _____ female horse
3. legion _____ large group of soldiers or people
4. mare _____ a paper that provides information
5. pulse
6. tub

1. concrete
2. era _____ circular shape
3. fiber _____ top of a mountain
4. loop _____ a long period of time
5. plank
6. summit

1. blend
2. devise _____ mix together
3. hug _____ plan or invent
4. lease _____ hold tightly in your arms
5. plague
6. reject

1. abolish
2. drip _____ bring to an end by law
3. insert _____ guess about the future
4. predict _____ calm or comfort someone
5. soothe
6. thrive

1. bleed
2. collapse _____ come before
3. precede _____ fall down suddenly
4. reject _____ move with quick steps and jumps
5. skip
6. tease

1. casual
2. desolate _____ sweet-smelling
3. fragrant _____ only one of its kind
4. radical _____ good for your health
5. unique
6. wholesome

1. gloomy
2. gross _____ empty
3. infinite _____ dark or sad
4. limp _____ without end
5. slim
6. vacant

Version 1: The 10,000-word level

1. antics
2. batch _____ foolish behavior
3. connoisseur _____ a group of things
4. foreboding _____ person with a good knowledge of art or music
5. haunch
6. scaffold

1. auspices
2. dregs _____ confused mixture
3. hostage _____ natural liquid present in the mouth
4. jumble _____ worst and most useless parts of anything
5. saliva
6. truce

1. casualty
2. flurry _____ someone killed or injured
3. froth _____ being away from other people
4. revelry _____ noisy and happy celebration
5. rut
6. seclusion

1. apparition
2. botany _____ ghost
3. expulsion _____ study of plants
4. insolence _____ small pool of water
5. leash
6. puddle

1. arsenal
2. barracks _____ happiness
3. deacon _____ difficult situation
4. felicity _____ minister in a church
5. predicament
6. spore

1. acquiesce
2. bask _____ to accept without protest
3. crease _____ sit or lie enjoying warmth
4. demolish _____ make a fold on cloth or paper
5. overhaul
6. rape

1. blaspheme
2. endorse _____ slip or slide
3. nurture _____ give care and food to
4. skid _____ speak badly about God
5. squint
6. straggle

1. clinch
2. jot _____ move very fast
3. mutilate _____ injure or damage
4. smolder _____ burn slowly without flame
5. topple
6. whiz

1. auxiliary
2. candid ___ bad-tempered
3. luscious ___ full of self-importance
4. morose ___ helping, adding support
5. pallid
6. pompous

1. dubious
2. impudent _____ rude
3. languid _____ very ancient
4. motley _____ of many different kinds
5. opaque
6. primeval

Version 1: Academic vocabulary

1. benefit
2. labor _____ work
3. percent _____ part of 100
4. principle _____ general idea used to guide one's actions
5. source
6. survey

1. element
2. fund _____ money for a special purpose
3. layer _____ skilled way of doing something
4. philosophy _____ study of the meaning of life
5. proportion
6. technique

1. consent
2. enforcement _____ total
3. investigation _____ agreement or permission
4. parameter _____ trying to find information about something
5. sum
6. trend

1. decade
2. fee _____ 10 years
3. file _____ subject of a discussion
4. incidence _____ money paid for services
5. perspective
6. topic

1. colleague
2. erosion _____ action against the law
3. format _____ wearing away gradually
4. inclination _____ shape or size of something
5. panel
6. violation

1. achieve
2. conceive _____ change
3. grant _____ connect together
4. link _____ finish successfully
5. modify
6. offset

1. convert
2. design _____ keep out
3. exclude _____ stay alive
4. facilitate _____ change from one thing into another
5. indicate
6. survive

1. anticipate
2. compile _____ control something skillfully
3. convince _____ expect something will happen
4. denote _____ produce books and newspapers
5. manipulate
6. publish

1. equivalent
2. financial _____ most important
3. forthcoming _____ concerning sight
4. primary _____ concerning money
5. random
6. visual

1. alternative
2. ambiguous _____ last or most important
3. empirical _____ something different that can be chosen
4. ethnic _____ concerning people from a certain nation
5. mutual
6. ultimate

Appendix 4: Experimental vs. control participants' scores on VLT

No	The experimental participants				The corresponding control subjects				The difference between the two overall scores
	Student ID	The 2000 WL score	Off-list score	Overall score	Student ID	The 2000 WL score	Off-list score	Overall score	
1	10515	16	15	31	11719	16	14	30	1
2	9044	22	18	40	9692	24	18	42	2
3	10736	26	18	44	12387	26	16	42	2
4	12067	23	20	43	9575	24	19	43	0
5	9211	30	16	46	12022	29	16	45	1
6	10373	24	24	48	9883	23	25	48	0
7	10782	27	28	55	10513	27	28	55	0
8	10997	20	9	29	11920	20	9	29	0
9	9851	20	13	33	10999	19	14	33	0
10	11288	26	36	62	9317	26	36	62	0
11	12839	28	31	59	11634	28	30	58	1
12	10115	25	25	50	10021	25	23	48	2
13	9219	21	17	38	10027	21	15	36	2
14	11042	24	13	37	9563	24	14	38	1
15	10281	16	10	26	9432	17	9	26	0
16	10299	25	15	40	12492	25	16	41	1
17	12001	18	2	20	11517	18	3	21	1
18	9162	26	28	54	10396	26	28	54	0
19	11111	14	3	17	12305	17	2	19	2
20	10943	22	15	37	10637	19	14	33	4
21	12702	28	18	46	12592	28	17	45	1
22	10472	13	11	24	10332	16	11	27	3
23	10394	26	18	44	11493	24	17	41	3
24	10662	19	10	29	11002	19	10	29	0
25	10974	26	14	40	11529	24	14	38	2
26	11279	27	19	46	10962	28	17	45	1
27	10882	23	20	43	10885	23	18	41	2
28	11602	20	14	34	11358	20	14	34	0
29	12120	18	12	30	10791	16	14	30	0
30	11003	16	15	31	10607	14	13	27	4
31	10989	19	11	30	12159	16	14	30	0
32	10729	22	11	33	11382	22	11	33	0
33	12729	27	20	47	10803	27	19	46	1
34	11863	30	44	74	11973	30	47	77	2
35	10547	18	4	22	12394	18	3	21	1
36	10648	26	14	40	12763	25	15	40	0
37	10739	28	27	55	11178	28	28	56	1
38	12499	24	17	41	10263	24	17	41	0
39	11735	22	13	35	11672	24	13	37	2
40	11531	16	2	18	11892	18	3	21	3

Appendix 5: List of the target Off-list words

abbey	barn	buoy	clumsy	defang	embarrassed	gadfly	helm	kitten	moat	pedestrian	quarry	scrimp	splash	tootle	yaw
abbot	barrow	burgle	clunker	defrost	emptiness	galaxy	helmet	knight	mobster	periscope	railhead	scrub	splenetic	torpid	yowl
aboard	basement	butler	coach	deliberate	emulated	gallop	herb	lantern	molar	perjure	ramp	sculptor	spool	tow	yucky
abortion	batch	cabin	cockpit	demolish	endurable	gamble	hijacker	latch	molest	petal	ranger	sculpture	squall	townie	
abracadabra	bedchamber	canoe	coffin	depot	enthusiasm	gang	hiss	leash	monastery	petition	ransom	scuttled	squat	tracker	
abscess	begrudge	canonize	colonel	depression	escort	gasp	homesick	legitimate	monastic	petticoat	rash	seasick	squeak	traitor	
adultery	bench	canteen	combat	desolate	excursion	gaudy	hood	leisure	monk	pilgrimage	ratify	segregate	quelch	tranquil	
affection	binge	canter	competent	desperate	exempt	genealogy	hoodwink	lettuce	monster	pistol	rebel	sergeant	squirrel	transport	
agitate	bingo	canvass	concert	despondent	exhaustion	gherkin	hoot	lizard	moor	pitilessly	recipe	sermon	squish	triumph	
alibi	binoculars	capricious	condemn	dialect	exotic	giant	hostess	loft	moral	pliable	redden	shed	stag	tug	
allergy	bishop	capsize	confidential	diarrhoea	expedite	giggle	hotbed	logo	morbid	poker	refugee	shepherd	stall	tutor	
ambassador	blackmail	caravan	console	dick	expel	girlish	hound	loopy	mould	pollute	rehearse	shoehorn	stare	vampire	
ancestry	blacksmith	cargo	constable	diddle	fabulous	gladius	howl	lounge	mourn	pond	relic	shrewd	statue	vandal	
anchor	blaze	carjacking	contagious	dinghy	fanatic	glassy	immunity	lumbering	mouthpiece	poppy	repose	shrine	stepmother	vase	
antique	blonde	carnival	cop	dire	fascinate	glue	imposter	lumberjack	mug	porch	rick	shrug	stern	vengeance	
aquarium	blunt	carp	coronation	discreet	faucet	gong	impress	magistrate	mule	pornography	riddle	sieve	stitch	veranda	
armour	blush	ceiling	corset	diverge	ferry	gosling	inattentive	magnet	nasty	porridge	riot	sigh	stride	vicar	
arsenal	boar	cellar	counsel	divorce	festival	gossip	indulge	maid	navigate	porter	robot	sin	stud	vicious	
artifice	bobble	cemetery	crawl	docket	festive	gramophone	infection	mangle	necropolis	pottery	roost	siren	stygian	violet	
assassinate	bog	chafe	cremate	dodgy	fetch	grange	inflict	mansion	neigh	pram	roulette	skate	sugarcane	vulgar	
athlete	bong	chamber	crew	doff	fiddle	gravestone	ingrate	marital	nibble	pregnancy	routine	skeleton	suicide	waddle	
atmosphere	bosom	chant	cricketer	dogma	fiord	graveyard	inherit	marshal	nightmare	premature	rump	skulduggery	summit	wade	
attitudinal	boulevard	chapel	crisp	domicile	flagon	grieve	inject	mast	noose	primitive	runway	slam	sump	wagon	
avalanche	bounty	chaplain	crook	dormitory	flagrant	grimly	innocent	mayor	obese	privilege	sapient	sledge	superstition	warble	
awesome	boycott	charitable	crow	doubter	flake	grin	inquest	maze	obscure	profit	savage	sleek	tank	warehouse	
awful	brash	chauffeur	crutch	dreary	flask	grouch	insistence	meagre	opera	prorate	scaffold	sleeve	tarmac	waterproof	
bacon	breed	cheep	curfew	drench	flicker	growl	inspire	merciless	ordain	protest	scandal	sleigh	telegram	whine	
bailiff	brew	chink	curtsy	duke	flute	grubby	interchange	midriff	orphanage	puck	scar	slipper	tempest	whoop	
balaclava	bricklayer	circus	dagger	dumb	forecast	hale	intertwine	midwife	outrage	punch	scarf	slosh	temptation	whoosh	
baldness	bridegroom	clan	damn	dumbly	forfeit	harness	invasion	miracle	paddle	puritan	scarlet	sob	tenant	wink	
ballroom	bridesmaid	clansman	dawn	dutiful	fortnight	harpoon	jiggle	misery	pantry	purity	scene	sorrowful	terrace	witch	
bandage	bruise	climate	debility	dwarf	fountain	haystack	jug	misfortune	parry	purse	scoundrel	spark	thrift	wodge	
bang	buggy	cloak	decant	echo	frown	headmaster	junction	mist	passion	pyjamas	scourge	sparkle	thumbscrew	wonky	
barbarian	bullet	cluck	deck	elegance	funky	hedge	kidnap	mistress	passionate	quaint	scratchy	species	toggle	yacht	
barge	bumble	clue	deface	elevator	furious	hellish	kite	moan	pasture	quake	scream	spectator	tomahawk	yam	

Appendix 6: List of the target 2K words

ambition	instant	strip	mineral	plough	burst
astonishment	ladder	stuff	murder	rob	mill
audience	madness	sympathy	oar	sore	pile
beggar	murderer	tempt	pattern	wreck	rush
bitterly	neat	thread	persuade	astonish	carriage
bow	pause	barrel	purple	bribe	pale
bunch	peculiar	boundary	shallow	bucket	pearl
bush	pinch	brick	shilling	cottage	confess
ceremony	pink	canal	slope	damp	cruel
companion	pot	cape	stretch	mysterious	decay
confident	pretend	cart	theatre	mystery	offend
cope	relieve	cattle	violent	spade	tremble
crack	reputation	chimney	wicked	avenue	hinder
curly	rescue	conscious	awkward	bargain	feast
custom	responsible	cultivate	bless	firm	
debt	ribbon	damper	bowl	widow	
delighted	ripe	ditch	burial	cliff	
edge	rival	drum	cautious	priest	
entertain	row	dull	fate	puzzle	
extreme	rub	explode	fierce	revenge	
fancy	rusty	fond	glory	scratch	
float	saddle	grace	harbour	steep	
generous	sew	hook	heap	steer	
handkerchief	spat	interfere	log	stiff	
hut	strap	mill	mend	trap	

Appendix 7: Sample of reading materials

The Silver Sword

'Perhaps she escaped, like you,' said Mrs Krause. 'Did you decide to meet somewhere if you were separated?'

'Yes, we did. In Switzerland. My wife is Swiss and her parents still live there.'

Mrs Krause smiled. 'Then go to Switzerland, and perhaps you will find her there.'

But Joseph spent several more days looking for his children. One afternoon, he was searching among the ruins of his old home when he found a small silver sword. It was about fourteen centimetres long, with a dragon at one end. It was a paper knife, used for opening letters. Joseph had once given it to his wife for a birthday present.

While he was cleaning the knife, he saw a small boy watching him. The boy was thin and his clothes were old and dirty. He was carrying a wooden box under one arm, and a small grey cat under the other.

'Give me that sword,' said the boy.

'But it's mine,' said Joseph.

'You found it here, and this is my place.'

Joseph explained about his house.

'I'll give you food for it,' said the boy, and he offered Joseph a sandwich.

'I have plenty of food,' said Joseph. He put his hand into his pocket, but it was empty. 'That's *my* sandwich!' he laughed. 'You took it from my pocket!'

But before Joseph could take it back, the boy ate most of it and gave the rest to his cat.

After a minute, Joseph said, 'I'm looking for my children.'

10

The Silver Sword

'I'm starting the journey to Switzerland tonight,' said Joseph. 'I'm going to hide on a train. Where's the best place to jump on a train unseen?'

'You will be caught and shot,' said the boy. 'Or you will die from the cold.'

'I still have to go,' said Joseph.

'Meet me tonight, when it's dark, and I'll show you the place where the trains slow down,' said the boy.

That night, when it was dark, Joseph said goodbye to the Krauses and left their house for the last time. The boy was waiting for him at the bottom of the street.

'We must use the back streets,' said the boy. 'If the Nazi soldiers see us, they'll shoot.'

'What's that you're carrying?' said Joseph.

'Bread,' said the boy. 'I borrowed it from the Nazi soldiers. They have plenty of it. Take it, you'll be hungry.'

'I've a lot to thank you for,' said Joseph, as they waited beside the railway. 'What's your name?'

The boy said nothing. He sat holding his cat and the wooden box.

'Will you come with me?' asked Joseph.

The boy didn't answer the question. He opened the wooden box and took out the silver sword. 'This will bring me luck, and it will bring you luck because you gave it to me. I don't tell anybody my name – it's not safe. But I'll tell you because you gave me the sword.' He whispered. 'It's Jan.'

A train was coming.

The silver sword



'Give me that sword,' said the boy.

Ruth is fifteen now, and she's tall with fair hair. Edek is thirteen, and Bronia is five.'

'Warsaw is full of children,' said the boy. 'They're all dirty and hungry and they all look alike.'

'I'll give you this sword if you do something for me,' said Joseph. 'If you ever see Ruth or Edek or Bronia, you must tell them about our meeting. Tell them I'm going to Switzerland to find their mother. Tell them to follow me as soon as they can.'

The boy took the sword and put it in his wooden box.

11

The children

'Goodbye, Jan,' said Joseph. 'Remember your promise. Whatever happens, I shall not forget you.'

It was dark, and Jan did not see him jump on to the train. It was raining heavily now, and Jan hurried back into the dark streets, with the grey cat inside his coat. The wooden box was under his arm.

And he thought of the silver sword inside.

4

The children

What happened to Joseph's family that night over a year ago? Was Mrs Krause's story true? Did the Nazi soldiers take Joseph's wife away? Did they return and blow up the house with the children in it?

This is what happened.

It was snowing that night in Warsaw. Ruth and Bronia were asleep in the room next to their mother's bedroom. Edek's room was on the top floor. He was asleep when the Nazi soldiers came, but woke up when he heard a noise outside his door.

The door was locked. Edek shouted and banged on it, but could not get out. He listened. In his mother's room, the men were giving orders, but Edek could not hear what they were saying. In the ceiling was a small square door that led to the attic. There was a ladder between his bed and the wall. Quietly, he moved it under the square door and climbed up.

There was a rifle hidden in the attic, and Edek took it and climbed back down to his room. The noises in the room below had stopped. He looked out into the street and saw a car waiting outside the front door. Two Nazi soldiers were pulling his mother towards it.

He opened the window. He was afraid to shoot until his mother was safely in the car. His first shot hit a soldier's arm. The man shouted with pain and jumped in beside the driver. Edek aimed the next two shots at the car wheels. He hit one, but the car got away.

Edek used the rifle to break down his bedroom door, then he did the same to the door of his sisters' room. Bronia was crying and Ruth was trying to calm her.

'I hit one of them,' said Edek.

'That was silly,' said Ruth. 'They'll come back for us now. We must get away from here before they do.'

Ruth dressed Bronia while Edek fetched overcoats and boots and warm caps. Ruth pulled a coat on over her nightdress, and put a scarf round Bronia.

'We can't go out the front way. I can hear another car coming,' said Edek. 'And the back wall is too high and there are soldiers in that street. We'll have to go over the roof.'

He picked up Bronia and led the way upstairs. He was wearing his father's thick overcoat and carrying the rifle on his back.

When they were in the attic, Edek broke the window to the roof and climbed out into the cold night. Ruth lifted Bronia up to him, then followed her.

'Listen, Bronia,' said Edek. 'If you make a sound, we shall all be killed. Walk behind me and hold on to the rifle. And don't look down!'

The roof was steep, and the snow made it difficult to stand or walk. Edek managed to climb across to the chimney, with Bronia holding on to the rifle behind him.



'Walk behind me and hold on to the rifle,' said Edek.

She was too afraid to speak or make a noise. Then he reached back and pulled Ruth up after him. They could not see what was happening in the street, but they could hear shouting and the sound of cars stopping suddenly.

The houses in this street were joined together, and so they were able to move from roof to roof and get away. They had gone a hundred metres when the first bomb exploded. Fire lit up the sky above their home, and they fell flat in the snow. The roof shook and the whole city seemed to tremble. Another bomb exploded, and smoke and flames came from the windows.

'Hurry,' said Edek. 'We won't let them get us now.'

They moved quickly across the roof-tops until they found a fire escape on the outside of a building, then they went down to the street. On they ran, not knowing or caring where they went as long as they left the terrible flames behind them.

It was the beginning of another grey winter's day before they finally stopped at a ruin of a bombed house. They slept inside it until the early afternoon, then woke up cold and hungry.

They made their new home in a cellar at the other end of the city. When they asked the Polish Council about their mother, they were told she had been taken to Germany to work on the land. Nobody knew which part of Germany.

'The war will end soon,' they were told, 'and your mother will come back.'

They quickly made their new home as comfortable as they could. Edek got a mattress and some curtains from a bombed building. He gave the mattress to Ruth and Bronia. The curtains made good sheets. He stole blankets from a Nazi camp, one for each of them. Here they lived for the rest of that winter and the spring.

Food was not easy to find. Except when Edek found work for a few days, there was no money to buy any. Sometimes they begged for it, other times they stole it from the Nazis. They saw nothing wrong in stealing from their enemies, but they were careful never to steal from their own people.



Edek got a mattress and some curtains from a bombed building.

Appendix 9: **Non-parametric Wilcoxon Signed Ranks Test**

NPART TESTS

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-2**. See **page 186** : Comparison of the experimental and control groups for the overall pre-vocabulary test scores assessed by means of the Paired-Samples t-Test

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Experimental.group.	18	0.5239	0.14026	0.3	0.71
Control.group	18	0.4732	0.13452	0.19	0.87

Wilcoxon Signed Ranks Test

Ranks

	N	Mean Rank	Sum of Ranks
Negative Ranks	10 ^a	11.6	116
Positive Ranks	8 ^b	6.88	55
Ties	0 ^c		
Total	18		

- a. Control.group.Overall.pre.test.average.score < Experimental.group.Overall.pre.test.average.score
- b. Control.group.Overall.pre.test.average.score > Experimental.group.Overall.pre.test.average.score
- c. Control.group.Overall.pre.test.average.score = Experimental.group.Overall.pre.test.average.score

Test Statistics^a

	Control.group - Experimental.group
Z	-1.328 ^b
Asymp. Sig. (2-tailed)	0.184

- a. Wilcoxon Signed Ranks Test
- b. Based on positive ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-3**. See **page 187**: Comparison of the experimental and control groups on pre-test off-list item scores, assessed by means of the Paired-Samples t-test.

	N	Mean	Std. Deviation	Minimum	Maximum
Experimental.group	18	0.4201	0.09879	0.26	0.56
Contreol.group	18	0.3688	0.15317	0.11	0.82

Wilcoxon Signed Ranks Test

	N	Mean Rank	Sum of Ranks
Negative Ranks	11 ^a	9.55	105
Positive Ranks	6 ^b	8	48
Ties	1 ^c		
Total	18		

- a. Contreol.group.pre.test.average.score.Offlist.words < Experimental.group.pre.test.average.score.Offlist.words
 b. Contreol.group.pre.test.average.score.Offlist.words > Experimental.group.pre.test.average.score.Offlist.words
 c. Contreol.group.pre.test.average.score.Offlist.words = Experimental.group.pre.test.average.score.Offlist.words

	Contreol.group- Experimental.group
Z	-1.349 ^b
Asymp. Sig. (2-tailed)	0.177

- a. Wilcoxon Signed Ranks Test
 b. Based on positive ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-4**. See **page 187**: Comparison of the experimental and control groups on pre-test 2K items scores, assessed by means of the Paired-Samples t-test

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Experimental.group	18	0.7366	0.3067	0.15	1.30
Control.group	18	0.6869	0.208	0.35	1.00

Wilcoxon Signed Ranks Test

Ranks

	N	Mean Rank	Sum of Ranks
Negative Ranks	12 ^a	8.33	100.00
Positive Ranks	6 ^b	11.83	71.00
Ties	0 ^c		
Total	18		

- a. Control.group.pre.test.average.score.2K < Experimental.group.pre.test.average.score.2K
- b. Control.group.pre.test.average.score.2K > Experimental.group.pre.test.average.score.2K
- c. Control.group.pre.test.average.score.2K = Experimental.group.pre.test.average.score.2K

Test Statistics^a

	Control.group - Experimental.group
Z	-.632 ^b
Asymp. Sig. (2-tailed)	0.527

- a. Wilcoxon Signed Ranks Test
- b. Based on positive ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-7**. See **page 191**: Comparison of the experimental and control group post-vocabulary test scores, assessed by means of the Paired Samples t-t-test.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Experimental.group.	18	1.1574	0.29662	0.63	1.60
Control.group.	18	0.4893	0.11257	0.25	0.74

Wilcoxon Signed Ranks Test

Ranks

	N	Mean Rank	Sum of Ranks
Negative Ranks	18 ^a	9.5	171
Control.group - Experimental.group. Positive Ranks	0 ^b	0	0.00
Ties	0 ^c		
Total	18		

- a. Control.group.Overall.first.post.test.average.score < Experimental.group.Overall.first.post.test.average.score
- b. Control.group.Overall.first.post.test.average.score > Experimental.group.Overall.first.post.test.average.score
- c. Control.group.Overall.first.post.test.average.score = Experimental.group.Overall.first.post.test.average.score

Test Statistics^a

	Control.group - Experimental.group
Z	-3.724 ^b
Asymp. Sig. (2-tailed)	0.00

- a. Wilcoxon Signed Ranks Test
- b. Based on positive ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-8**. See **page 192**: Mean comparison of the experimental and control groups on the pre- and post-test

Descriptive Statistics					
	N	Mean	Std. Deviation	Minimum	Maximum
Experimental.group (pre-test)	18	0.5239	0.14026	0.30	0.71
Control.group.(pre-test)	18	0.4732	0.13452	0.19	0.87
Experimental.group(post-test)	18	1.1574	0.29662	0.63	1.60
Control.group(post-test)	18	0.4893	0.11257	0.25	0.74

Wilcoxon Signed Ranks Test

Ranks			
	N	Mean Rank	Sum of Ranks
Negative Ranks	0 ^a	0	0.00
Experimental.group(.post.test)- Experimental.group(pre.test)	18 ^b	9.5	171.00
Ties	0 ^c		
Total	18		
Negative Ranks	7 ^d	7.57	53.00
Control.group(.post.test)- Control.group(pre.test)	10 ^e	10	100.00
Ties	1 ^f		
Total	18		

- a. Experimental.group (post.test) < Experimental.group (pre.test)
 b. Experimental.group(post.test) > Experimental.group(pre.test)
 c. Experimental.group(post.test) = Experimental.group(pre.test)
 d. Control.group(post.test) < Control.group(pre.test)
 e. Control.group(post.test) > Control.group(pre.test)
 f. Control.group(post.test) = Control.group(pre.test)

Test Statistics ^a		
	Experimental.group (.post.test)- Experimental.group (pre.test)	Control.group (.post.test) - Control.group (pre.test)
Z	-3.724 ^b	-1.112 ^b
Asymp. Sig. (2-tailed)	0.00	0.27

- a. Wilcoxon Signed Ranks Test
 b. Based on positive ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-9**. See **page 193**: Mean percentage difference for words rated by the experimental group at each the four word knowledge scale levels in the pre-and post-test.

	N	Mean	Std. Deviation	Minimum	Maximum
No knowledge. pre-test.	18	62.0072	5.58564	50.67	71.70
Slight knowledge. Pre-test ..	18	27.8921	5.26501	20.75	38.67
Receptive knowledge. pre-test	18	5.8013	3.44392	.00	12.82
Productive knowledge. pre-test	18	4.2994	4.15164	.00	13.33
No knowledge. post-test.	18	35.0205	6.99587	20.75	46.88
Slight knowledge. Post-test ..	18	30.8436	13.02274	13.08	58.49
Receptive knowledge. post-test	18	17.5119	7.41445	3.70	33.64
Productive knowledge. post-test	18	16.6240	10.76689	.00	38.04

Wilcoxon Signed Ranks Test

		N	Mean Rank	Sum of Ranks
No knowledge. post-test - No knowledge. pre-test	Negative Ranks	18 ^a	9.50	171.00
	Positive Ranks	0 ^b	.00	.00
	Ties	0 ^c		
	Total	18		
Slight knowledge. Post-test - Slight knowledge. Pre-test	Negative Ranks	7 ^d	7.71	54.00
	Positive Ranks	9 ^e	9.11	82.00
	Ties	2 ^f		
Receptive knowledge. post-test - Receptive knowledge. pre-test	Negative Ranks	0 ^g	.00	.00
	Positive Ranks	18 ^h	9.50	171.00
	Ties	0 ⁱ		
Productive knowledge. post- Productive knowledge. pre-test	Negative Ranks	0 ⁱ	.00	.00
	Positive Ranks	15 ^k	8.00	120.00
	Ties	3 ^l		
Total		18		

	No knowledge. post-test - No knowledge. pre-test	Slight knowledge. Post-test - Slight knowledge. Pre-test	Receptive knowledge. post-test - Receptive knowledge. pre-test	Productive knowledge. post- Productive knowledge. pre-test
Z	-3.724 ^b	-.724 ^c	-3.724 ^c	-3.408 ^c
Asymp. Sig. (2-tailed)	.000	.469	.000	.001

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-10**. See **page 195**: Mean percentage difference of words rated by the control group at each of the four word knowledge scale levels in the pre-and post-test.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
No knowledge. pre-test.	18	67.4407	8.32779	43.48	84.91
Slight knowledge. Pre-test .	18	21.5604	6.92766	8.97	33.70
Receptive knowledge. pre-test	18	7.2401	3.62847	2.13	15.22
Productive knowledge. pre-test	18	3.7587	3.27576	.00	10.00
No knowledge. post-test.	18	64.7313	5.64768	55.43	77.36
Slight knowledge. Post-test ..	18	25.0840	4.08039	18.67	32.50
Receptive knowledge. post-test	18	6.7047	3.07759	1.89	14.13
Productive knowledge. post-test	18	3.4801	2.77034	.00	8.33

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
No knowledge. post-test - No knowledge. pre-test	Negative Ranks	12 ^a	8.67	104.00
	Positive Ranks	4 ^b	8.00	32.00
	Ties	2 ^c		
	Total	18		
Slight knowledge. Post-test - Slight knowledge. Pre-test	Negative Ranks	4 ^d	7.75	31.00
	Positive Ranks	13 ^e	9.38	122.00
	Ties	1 ^f		
Receptive knowledge. post-test - Receptive knowledge. pre-test	Total	18		
	Negative Ranks	7 ^g	5.14	36.00
	Positive Ranks	2 ^h	4.50	9.00
Productive knowledge. post-Productive knowledge. pre-test	Ties	9 ⁱ		
	Total	18		
	Negative Ranks	6 ^j	3.67	22.00
	Positive Ranks	2 ^k	7.00	14.00
	Ties	10 ^l		
	Total	18		

Test Statistics^a

	No knowledge. post-test - No knowledge. pre-test	Slight knowledge. Post-test - Slight knowledge. Pre-test	Receptive knowledge. post-test - Receptive knowledge. pre-test	Productive knowledge. post-Productive knowledge. pre-test
Z	-1.862 ^b	-2.154 ^c	-1.601 ^b	-.561 ^b
Asymp. Sig. (2-tailed)	.063	.031	.109	.575

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-12**. See **page 200**: Comparison of the experimental and control group for post-test off-list item scores, assessed by means of the Paired-Samples t-test

	N	Mean	Std. Deviation	Minimum	Maximum
Experimental.group(.post.test.Offlist.words)	18	1.1249	0.28866	0.63	1.60
Control.group.first(post.test.Offlist.words)	18	0.3682	0.12726	0.19	0.73

Wilcoxon Signed Ranks Test

	N	Mean Rank	Sum of Ranks
Negative Ranks Control.group(post.test.Offlist.words) - Experimental.group(post.test.Offlist.words)	18 ^a	9.50	171
Positive Ranks	0 ^b	0.00	0.00
Ties	0 ^c		
Total	18		

- a. Control.group(.post.test. Offlist.words) < Experimental.group(.post.test. Offlist.words)
 b. Control.group(.post.test. Offlist.words) < > Experimental.group(.post.test. Offlist.words)
 c. Control.group(.post.test. Offlist.words) < = Experimental.group(.post.test. Offlist.words)

	Control.group.(post.test.Offlist.words) - Experimental.group(post.test.Offlist.words)
Z	-3.724 ^b
Asymp. Sig. (2-tailed)	0.00

- a. Wilcoxon Signed Ranks Test
 b. Based on positive ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-13**. See **page 201**: Comparison of the experimental and control group levels of knowledge of off-list items in the pre-and post-test

Wilcoxon Signed Ranks Test

	Test Statistics ^a Experimental group's offlist items (pre-post tests)			
	Perc.Post1.offlist.chice1.Exp - Perc.offlistpre.choice1.Exp	Perc.Post1.offlist.chice2.Exp - Perc.offlistpre.choice2.Exp	Perc.Post1.offlist.chice3.Exp - Perc.offlistpre.choice3.Exp	Perc.Post1.offlist.chice4.Exp - Perc.offlistpre.choice4.Exp
Z	-3.724 ^b	-1.207 ^c	-3.621 ^c	-3.408 ^c
Asymp. Sig. (2-tailed)	.000	.227	.000	.001

- a. Wilcoxon Signed Ranks Test
- b. Based on positive ranks.
- c. Based on negative ranks.

	Test Statistics ^a Control group's offlist items (pre-post tests)			
	Perc.Post1.offlist.chice1.Control grp - Perc.offlistpre.choice1.Control grp	Perc.Post1.offlist.chice2.control grp - Perc.offlistpre.choice2.control grp	Perc.Post1.offlist.chice3.control grp - Perc.offlistpre.choice3.control grp	Perc.Post1.offlist.chice4.conm trol grp - Perc.offlistpre.choice4.control grp
Z	-.646 ^b	-1.254 ^c	-1.956 ^b	-.135 ^b
Asymp. Sig. (2-tailed)	.518	.210	.050	.893

- a. Wilcoxon Signed Ranks Test
- b. Based on positive ranks.
- c. Based on negative ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-15**. See **page 204**: Comparison of the experimental and control group on post-test off-list items scores assessed by means of the Paired-Samples t-Test

	N	Mean	Std. Deviation	Minimum	Maximum
Experimental.group.(post.test.2K.words)	18	1.2236	0.38058	0.56	2.10
Control.group.(post.test.2K.words)	18	0.7373	0.18851	0.35	1.00

Wilcoxon Signed Ranks Test

	N	Mean Rank	Sum of Ranks
Negative Ranks	16 ^a	10.38	166
Control.group.(post.test..2K.words) Experimental.group.(post.test..2K.words)	-		
Positive Ranks	2 ^b	2.50	5.00
Ties	0 ^c		
Total	18		

a. Control.group.(post.test. 2K.words) < Experimental.group. .(post.test. 2K.words)

b. Control.group. .(post.test. 2K.words) > Experimental.group.(post.test. 2K.words)

c. Control.group.(post.test. 2K.words) = Experimental.group.(post.test. 2K.words)

	Control.group.(post.test.2K.words) - Experimental.group.(post.test.2K.words)
Z	-3.506 ^b
Asymp. Sig. (2-tailed)	0.00

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Non-parametric Wilcoxon Signed Ranks Test output equivalents for results in **Table 6-16**. See **page 205**: Comparison of the experimental and control group levels of knowledge 2K items in the pre-and post-test

Wilcoxon Signed Ranks Test

	Test Statistics ^a Experimental group's 2K items (pre-post tests)			
	Perc.Post1.2K.choice1.Exp - Perc.2Kpre.choice1.Exp	Perc.Post1.2K.choice2.Exp - Perc.2Kpre.choice2.Exp	Perc.Post1.2K.choice3.Exp - Perc.2Kpre.choice3.Exp	Perc.Post1.2K.choice4.Exp - Perc.2Kpre.choice4.Exp
Z	-3.362 ^b	-.852 ^b	-3.078 ^c	-3.412 ^c
Asymp. Sig. (2-tailed)	.001	.394	.002	.001

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

	Test Statistics ^a Control group's 2K items (pre-post tests)			
	Cont.Perc.2KPost1.Choice.1 - Cont.Perc.2K.Pre.Choice.1	Cont.Perc.2KPost1.Choice.2 - Cont.Perc.2K.Pre.Choice.2	Cont.Perc.2KPost1.Choice.3 - Cont.Perc.2K.Pre.Choice.3	Cont.Perc.2KPost1.Choice.4 - Cont.Perc.2K.Pre.Choice.4
Z	-2.274 ^b	-2.086 ^c	-.314 ^c	-.946 ^b
Asymp. Sig. (2-tailed)	.023	.037	.753	.344

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

GLOSSARY

- brake** (*n*) something that you use to stop a moving train or car
- charity** help or money for people who are poor or in trouble
- fireman** a man who keeps the fire burning in a steam engine
- government** the group of people who control a country
- hare** a small wild animal like a rabbit, with long ears
- horrible** very bad
- hound** a kind of dog which chases and catches wild animals like hares
- magic** something that makes strange and wonderful things happen, which nobody can explain
- maid** a woman who works in another person's house
- paint** (*n*) something wet and coloured which you use to make pictures or to change the colour of something
- petticoat** a kind of 'skirt' that a woman or girl wears under her dress
- platform** the part of a railway station where you stand to wait for a train
- porter** a person who works on a railway station and who carries suitcases
- pram** a kind of box on wheels to carry a baby in
- steam** a kind of 'smoke' which comes from very hot water
- wave** (*v*) to move a hand (or handkerchief) from one side to the other
- yard** a piece of hard ground near a building, usually with a wall round it