

**HONG KONG CHINESE SECONDARY SCHOOL
ENGLISH LANGUAGE LEARNERS' VOCABULARY
UPTAKE USING SIMPLIFIED READERS WITH AND
WITHOUT ASSOCIATED READING TASKS**

GRANT ANDREW DYKES

A DISSERTATION SUBMITTED TO LEICESTER UNIVERSITY

**in Partial Fulfilment of the Requirements for the Degree of
DOCTOR OF EDUCATION**

September 2011

HONG KONG CHINESE SECONDARY SCHOOL ENGLISH LANGUAGE LEARNERS' VOCABULARY UPTAKE USING SIMPLIFIED READERS WITH AND WITHOUT ASSOCIATED READING TASKS

GRANT A. DYKES

ABSTRACT

While there is a suggestion that reading extensively may be effective in improving vocabulary knowledge, a number of Hong Kong (H.K.) studies suggest that it has not been particularly successful in the H.K. context, particularly in aiding weaker language learners. The failure of extensive reading (ER) in H.K. could be attributable to a clash between the philosophical underpinnings of ER and the pedagogy and culture of the H.K. education system leading to faulty implementation of, and engagement with ER. Alternatively, the failure of ER in H.K. could be due to an inability of weaker L2 Chinese students to acquire language implicitly from uninstructed reading (UR).

The purpose of this study was to determine, by means of a quasi-experiment, whether L2 Chinese students could acquire language (vocabulary) implicitly from uninstructed reading and whether UR was suitable for language acquisition for both weaker and stronger students. The study also sought to determine whether UR was more or less effective than an alternative reading method, reading with tasks (R+), which employs implicit and explicit means of acquisition and may be more appropriate in the H.K. education system than UR in the form of ER.

The results revealed that both weaker and stronger readers could acquire vocabulary from UR, with little difference in acquisition between them. Both groups acquired significantly more vocabulary from R+ than from UR, although there was no advantage for either group. Although R+ required more time, vocabulary acquisition was more guaranteed than from UR. The key element of vocabulary acquisition from reading is text at an appropriate level and the study suggests that this may be closer to 99% comprehensibility than the 95% - 98% suggested by some researchers. R+ can be employed with a whole class, allowing both weaker and stronger readers to acquire a significant number of target words from a text. The ability of R+ to make a text easier, aid enjoyment, ensure interaction with text, fit the H.K. education system and scaffold ER is discussed and further research is suggested.

ACKNOWLEDGEMENTS

I would like to express my appreciation to Warwick B. Elley, Professor Emeritus of Canterbury University (NZ) who inspired me as an undergraduate in the 1980s with lectures on his research in Fiji and the power of reading. His work is still inspirational and I hope that in some small way I may have contributed to the field he pioneered.

Grateful thanks to my principal Ms. Lee Suet Ying for permission to carry out the study and to the students of the college for participating and allowing me to present their results.

I am indebted to MacMillan and Oxford University Press for permission to copy their excellent books for this research.

Grateful thanks to Dr. Kevin Armstrong who provided statistical insights and valuable feedback at all stages of this project.

Not a single word would have been written if it was not for the support of Delia Sagolili who had more faith in me than I had in myself and who accompanied me on this long journey. Thanks to Nicholas, Benjamin and Jacob (who was born midway through) who had little idea of what I was up to but taught me a great deal about language acquisition and demonstrated so much patience when I ‘just needed to finish something’. It is finished now guys!

TABLE OF CONTENTS

ABSTRACT.....	i
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS.....	iii
LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
GLOSSARY	xii

CHAPTER 1 INTRODUCTION **1**

1.0 Overview	1
1.1 Background	5
1.2 Terminology	8
1.3 The power of reading	10
1.3.1 Gains from reading	11
1.4 The introduction of ER in H.K.	15
1.5 Failure of extensive reading in H.K.	17
1.6 An alternative to extensive reading (ER)	24
1.7 Study outline	28

CHAPTER 2 REVIEW OF THE LITERATURE **30**

2.0 Introduction	30
2.1.0 Vocabulary acquisition from uninstructed reading (Part 1).....	33
2.1.1 The importance of vocabulary	33

TABLE OF CONTENTS (continued)

2.1.2	Implementation problems of extensive reading	37
2.1.3	Failure to acquire language from uninstructed reading	43
2.1.4	Summary of Part 1	57
2.2.0	Language theories of uninstructed reading (Part 2)	58
2.2.1	Implicit and explicit learning	58
2.2.2	The Noticing Hypothesis and key assumptions of UR	62
2.2.3	Opposition to the Noticing Hypothesis	75
2.2.4	Implicit learning: Krashen's Comprehension Hypothesis.....	81
2.2.5	Extensive reading studies	95
2.2.6	Summary of Part 2	102
2.3.0	Reading with tasks (Part 3)	104
2.3.1	Aims of R+	111
2.3.2	R+ as interactive reading	114
2.3.3	R+ and Schema development	117
2.3.4	Tasks and the task load hypothesis	119
2.3.5	Studies involving reading with tasks	128
2.4.0	Conclusion	146

CHAPTER 3 METHODOLOGY 149

3.0	Introduction	149
3.1.0	Methodology	149
3.1.1	Participants	155
3.1.2	Materials: Placement tests	158

TABLE OF CONTENTS (continued)

3.1.3	Materials: Book Selection	167
3.1.4	Materials: Worksheets and literature circle	185
3.1.5	Materials: Pretests and posttests	188
3.2.0	Research Procedures	191
3.2.1	General design	191
3.2.2	Quasi-experiment.....	194
3.3.0	Analysis	201
3.4.0	Considerations	202
3.4.1	Ethical considerations	202
3.4.2	Considerations of validity	203

CHAPTER 4 RESULTS AND ANALYSIS 205

4.0	Introduction.....	205
4.1.0	Results analysis (RA) for research question 1	208
4.1.1	RA: Weaker readers after 14 days.....	208
4.1.2	RA: Weaker readers after 28 days	209
4.1.3	RA: Stronger readers after 14 days	210
4.1.4	RA: Stronger readers after 28 days	210
4.2.0	Results analysis (RA) for research question 2	214
4.2.1	RA: Weaker readers after 14 days (R+)	215
4.2.2	RA: Weaker readers after 28 days (R+)	215
4.2.3	RA: Stronger readers after 14 days (R+).....	215
4.2.4	RA: Stronger readers after 28 days (R+)	216

TABLE OF CONTENTS (continued)

4.2.5 RA: All students after 14 days (R+).....	216
4.2.6 Summary and comparison tables for research questions 1 and 2	216
4.3.0 RA for research question 3	222
4.3.1 RA: Weaker readers R+ and UR after 14 days	223
4.3.2 RA: Weaker readers R+ and UR after 28 days	223
4.3.3 RA: Stronger readers R+ and UR after 14 day	223
4.3.4 RA: Stronger readers R+ and UR after 28 days	223
4.3.5 Group comparisons	224
4.3.6 RA: Vocabulary decay for UR	225
4.3.7 RA: Vocabulary decay across methods (WR).....	226
4.3.8 RA: Vocabulary decay across methods (SR)	227
4.3.9 RA: Vocabulary decay between SR and WR on R+	227
4.4.0 RA for research question 4	228
4.4.1 RA for survey questions	229
4.4.2 RA for survey question 12: Reasons for not reading	232
4.5.0 RA for research question 5	234
4.5.1 RA: WR and SR compared on uninstructed reading (14 days)	236
4.5.2 RA: WR and SR compared on uninstructed reading (14 days)	236
4.5.3 RA: WR and SR compared on reading with tasks (28 days)	237
4.5.4 RA: WR and SR compared on reading with tasks (28 days)	237
4.6.0 Summary of results	238

TABLE OF CONTENTS (continued)

CHAPTER 5 DISCUSSION	241
5.0 Introduction.....	241
5.1.0 Research question 1.....	241
5.1.1 Comparison of book difficulty and enjoyment for UR14.....	246
5.1.2 Comparison of book enjoyment and difficulty for UR28.....	247
5.1.3 Assumptions of UR.....	248
5.1.4 Motivation to read.....	256
5.1.5 Comprehensible reading level.....	260
5.2 Research question 2.....	265
5.3 Research question 3.....	270
5.4 Research questions 4 and 5.....	279
5.5 Limitations.....	283
 CHAPTER 6 CONCLUSION	 287
 REFERENCES	 296
 APPENDIXES – Attached CD ROM (PDF)	 324
Appendix A Tests, surveys and target word lists	CD 1
A1: Vocabulary placement tests – test A: levels test	1
A2: Vocabulary placement tests – test B: pat vocabulary	7
A3: Vocabulary placement tests – test C: pat comprehension	8

TABLE OF CONTENTS (continued)

A4: Vocabulary placement tests – test D: Nation and Beglar vocabulary test	9
A5: Selection test of the 26 target words	14
A6: Target word lists	16
A7: Pretests and posttests	20
A8: Cohen’s formula and effect size conversion table	29
A9: Reading survey	31
Appendix B Background information and studies	CD 34
B1: CMI / EMI Schools	34
B2: An explanation of the shared book approach	35
B3. Studies of reading with tasks (R+)	36
B4. Studies of extensive reading or uninstructed reading	37
Appendix C Tasks and worksheets	CD 39
C1: Literature circles	39
C2: Uninstructed book worksheets	41
C3: Reading with tasks worksheets	45
Appendix D: Letters of consent	CD 61
D1: Permission letter – Oxford University Press	61
D.2: Permission letter – MacMillan Readers	63
D.3: Student consent	64
D.4: School consent	65

LIST OF TABLES

Table 2.1	Summary of studies of vocabulary acquisition from reading	71
Table 2.2	Study duration and effect of ER on language acquisition, Krashen (2004b:2)	92
Table 2.3	A summary of the basic characteristics of ER, R+ and IR.....	113
Table 2.4	First and second semester gains compared (Smith 2006:14)	137
Table 2.5	CSEPT test results	138
Table 3.1	Response table for the number of books students had read in English	157
Table 3.2	Passes at each vocabulary level for test A	161
Table 3.3	Pearson Correlation of four placement tests	164
Table 3.4	Number of students passing at each vocabulary level and average scores	166
Table 3.5	Analysis of words, difficulty and enjoyment for selected books	174
Table 3.6	Number of occurrences of each target word	176
Table 3.7	Word frequencies comparing UR and R+ books for weaker readers	178
Table 3.8	Word frequencies comparing UR and R+ books for stronger readers	180
Table 3.9	Word frequencies comparing WR with SR	182
Table 3.10	Word frequencies comparing books between groups WR and SR for reading with tasks.....	184
Table 3.11	Two test sets from the pre / posttest for The Promise (WR UR28)	190
Table 3.12	Schedule for implementation of quasi-experiment	200
Table 4.1	Vocabulary gains from uninstructed Reading	212

LIST OF TABLES (continued)

Table 4.2	Number of students failing to acquire vocabulary from uninstructed reading.....	213
Table 4.3	Vocabulary gains for reading with tasks	217
Table 4.4	Effect sizes for research questions 1 and 2.....	218
Table 4.5	Failure to acquire vocabulary from R+	219
Table 4.6	Comparison of UR with R+	220
Table 4.7	Listed in ascending order of percentage gain from total words available.....	221
Table 4.8	ANOVA post-hoc test effect size summary for RA 4.4.0 to 4.4.4	224
Table 4.9	Survey questions with tallied Likert scores	230
Table 4.10	Reasons for not reading.....	233
Table 4.11	ANCOVA pre- and posttest mean scores.....	238
Table 5.1	Reasons for not reading with possible reasons for not reading the <i>first book</i> highlighted	258

LIST OF FIGURES

Figure 4.1	Vocabulary acquisition for SR and WR for UR and R+ measured after 14 and 28 days.....	184
------------	--	-----

GLOSSARY

AWL	ACADEMIC WORD LIST
CI	CONFIDENCE INTERVAL
CMI	CHINESE AS THE MEDIUM OF INSTRUCTION
EFL	ENGLISH AS A FOREIGN LANGUAGE
EMI	ENGLISH AS THE MEDIUM OF INSTRUCTION
ER	EXTENSIVE READING
HK	HONG KONG
IR	INTENSIVE READING
L1	FIRST LANGUAGE
L2	SECOND LANGUAGE
<i>M</i>	MEAN
<i>r</i>	PEARSON'S CORRELATION COEFFICIENT (USED AS AN EFFECT SIZE)
R+	READING WITH TASKS
RA	RESULTS ANALYSIS
S1	SECONDARY ONE
<i>SD</i>	STANDARD DEVIATION
SLA	SECOND LANGUAGE ACQUISITION
SR	STRONGER READERS
UR	UNINSTRUCTED READING
VC	VOCABULARY
WR	WEAKER READERS

CHAPTER 1

INTRODUCTION

1.0 Overview

A number of writers have suggested that exposure to a large quantity of text through extensive reading may enable language acquisition. Research (1.3) suggests that reading extensively can improve vocabulary knowledge, spelling, reading comprehension, reading speed and writing ability while providing knowledge of the world and improving learners' motivation to read. While extensive reading (ER) has been particularly popular in the West for students reading in their first language (L1), the last 20 years have seen an increase in its use for second language (L2) learners. It is seen as a relatively simple way to increase exposure to the target language in the belief that increased exposure will aid acquisition of the language. Providing second language students with books to read may be the only practical option for out-of-class language development for some learners in places such as Hong Kong (H.K.) where English is not commonly used.

Reading without instruction (uninstructed reading – UR) in the form of ER is not just a 'method' or 'technique' that can be easily slotted into a school programme, but encompasses an entire educational philosophy. It accepts that language acquisition is

possible without instruction and provides time for the implementation of the programme.

The ER philosophy embraces student-centered learning by allowing time for students to self-select books and to read at their own pace for pleasure. Many of the fundamental principles for ER are in stark contrast to the traditional H.K. educational philosophy where the teachers' role is to instruct students and then test what has been learnt. Local students have little experience with ER in their own language, making ER in English a foreign concept for many students, teachers and parents.

An extensive reading scheme was introduced into H.K. schools in the 1990s in an effort to raise English standards. It is suggested that the scheme has not greatly improved the language ability of the students involved (Wong, 2001), with students demonstrating few gains in controlled studies (Lai, 1991, 1993a, 1993b).

The lack of success of ER in H.K. could be due to a failure to fully embrace the guiding ER principles, leading to a lack of proper implementation of the programme. Alternatively, the failure of ER could be due to the inability of H.K. secondary school students to acquire language implicitly from uninstructed reading. The first question that this study seeks to answer, therefore, is whether a group of L2 H.K. secondary school students can acquire language (vocabulary) from uninstructed reading.

There is a suggestion in the literature that students with a ‘certain level of ability’ in English can learn to read by extensive reading alone by acquiring language implicitly as they read, although it is unclear what the ‘certain level’ of ability is. Controlled studies of ER in H.K. suggest that some students may benefit from ER some of the time. The second question this study seeks to answer is whether both weaker and stronger students can acquire language (vocabulary) from uninstructed reading.

The inherent philosophical difference between ER and the traditional H.K. system may mean that ER is not a particularly suitable method for language acquisition in the Hong Kong (H.K.) context. Some researchers have suggested that reading with tasks (R+) (implicit and explicit acquisition) may be more effective for language acquisition than uninstructed extensive reading (implicit acquisition). R+ may be able to offer learning assistance to students at all ability levels, particularly weaker students who may require it the most. In addition, R+ may fit the H.K. context better than ER by providing a method with which students are more familiar than that of simply reading for pleasure, as in uninstructed ER. The third research question this study seeks to address is which method (uninstructed reading – UR or reading with tasks – R+) leads to greater language (vocabulary) acquisition. As an extension to research questions one and two, question three compares the effectiveness of the two methods (UR and R+) to determine

whether one method is more effective than the other for a particular ability group. In other words, can stronger readers (SR) acquire more vocabulary from UR than weaker students (WR) or is R+ more suitable for one group or the other?

Very few studies have attempted to compare language acquisition for uninstructed reading (UR) and reading with tasks (R+), and no studies (to date) have examined whether one programme (UR or R+) is more effective for stronger or weaker readers. If one is to set up a school or nationwide ER programme, it is vital that all participants at all ability levels are able to benefit.

The purpose of this quasi-experimental study is to clarify whether Hong Kong school-aged second language (L2) learners can acquire language (vocabulary) from uninstructed reading (UR) and whether this method is more or less effective than reading with tasks (R+). The study also seeks to determine whether either method (UR or R+) is better suited for weaker or stronger L2 students for the purpose of acquiring vocabulary.

1.1 Background

Reading for pleasure as a means to acquire language is not a new concept in either first (L1) or second (L2) language learning. As a child attending primary school in New Zealand in the 1960s, we would spend a portion of every school day reading for pleasure (L1). Many years later, as a primary teacher, I ensured that time was available in my classes for pleasure reading (also called independent reading, extensive reading or sustained silent reading, SSR). Extensive reading or pleasure reading was a part of the curriculum in New Zealand primary schools from at least as far back as the 1960s and is still an important part of the New Zealand primary school curriculum today.

In earlier references to ER, Pilgreen (2000, pg. 1 - 2) does not exclusively use the term ER, but mentions the "Roots of SSR" going back to individualized reading or personalized reading programmes in elementary schools in the 1950s, most likely in the US. West (1932) uses the terms 'extensive method' (pg. 101) and 'extensive reading' (pg. 111).

Bond (1926) describes and analyses the extensive reading component of a French course in a Junior College where voluntary, informal, outside reading was added to the formal, assigned reading for classroom analysis. Results of extensive reading included a

trebling of the reading rate and a strong correlation between the amount of reading and comprehension and general achievement.

Hagboldt (1925), in the first year German course in the Junior College of the University of Chicago, not only employed extensive reading alongside intensive reading and the other usual features of a modern language course but also shifted the emphasis of the course completely away from grammar, placing it on reading instead. He suggested that the effect of extensive reading could not be replaced by any known means in modern language instruction (pg. 295).

The British writer / biographer Lucy Aikin produced graded versions of *Robinson Crusoe*, *Swiss Family Robinson*, *Aesop's Fables* and other classics *In Words of One Syllable* written in the 1850s / 60s for young readers, presumably to enable them to read the books for pleasure, and writing in the 1640s - 60s, the pioneer Czech educator John Comenius advocated graded reading, which was a learner-centered pedagogy (Monroe, 1900).

An early study introducing extensive reading to L2 students on the small Pacific island of Niue in the 1970s (De'Ath, 2001) paved the way for some larger-scale L2 studies in

later years. Elley and Mangubhai (1981) carried out what was to become the first of a series of large-scale research studies on the benefits of extensive reading on students in Fiji, involving more than 300 students. Later years saw similar large-scale studies carried out in a number of countries, including Singapore, Brunei, Sri Lanka and South Africa.

In more recent years, second language (L2) English extensive reading (ER) programmes have been implemented in Malawi in 1995 (Williams, 2007), Ethiopia (Ambatchew, 2004) and Cameroon (Davis, 1995; Tup and Shu, 1997) and have seen increasing popularity in Japan (Day et al, 1991).

The development of reading in a foreign language has become increasingly popular in second language circles, with discussion groups and online forums (<http://www.extensivereading.net/>) and entire journals dedicated to the subject (Reading in a Foreign Language - <http://nflrc.hawaii.edu/rfl/>).

Extensive reading was formally introduced into the H.K. school curriculum in the 1990s as a part of the English language curriculum to boost students' language ability.

1.2 Terminology

Our working definition of ‘extensive reading’ (ER) as a language teaching/learning procedure is that it is reading of (a) large quantities of material or long texts; (b) for global or general understanding; (c) with the intention of obtaining pleasure from the text. Further, because (d) reading is individualized, with students choosing the books they want to read, (e) the books are not discussed in class (Susser and Robb, 1990). ER is therefore uninstructed reading (UR) whereby readers acquire language through extensive exposure to it at an appropriate level.

A very large number of acronyms have been coined for extensive reading, largely in order to provide students with fun, catchy phrases such as SSR (sustained silent reading), DEAR (drop everything and read) and SURF, introduced in Hawaii (silent uninterrupted reading for fun). The extensive reading website discussion board (<http://www.extensivereading.net/>) lists some 20 acronyms. During these variously named reading sessions, the teacher may carry out brief conferencing with students to discuss what they have read, teachers may silently read themselves in order to model ‘correct practice’ and students may complete very brief reports or records on what they have read. The primary purpose of ER is to get students reading in the second language and liking it (Day and Bamford, 2000) and the key to achieving this, according to Grabe

and Stoller (2002:259), is to ensure that the material is 'within the students' linguistic competence'. Krashen (1989) suggests that input that is at a level whereby it is comprehended, such as that supplied through extensive reading, will naturally lead to acquisition. The purpose of ER is for students to gain pleasure from reading. Students self-select books, and where the text is at an appropriate level, they experience success, and this success helps to motivate students to want to read more (Day and Bamford, 2002).

There is some debate over exactly what large quantities of material or long texts actually comprise in order to be 'extensive'. Nation and Wang (1999) suggest a book a week, while Brown (2000) and Carroll (1972) suggest two graded readers a week. A great deal depends on the ability of the reader, as an advanced learner may only complete a single book in a fortnight, but the book may contain 40,000 words, as compared to a beginner reader completing three books a fortnight of only 1,500 words each. Hunt and Beglar (2005) comment that from their experience, less motivated and less proficient students found 40 pages in two weeks demanding. This measure, however, is rather dependent on the size of the print and the number of pictures included. For ER to be effective, regular exposure to text is required, such that daily reading and the completion of a book throughout a week or so is possibly more effective

than reading an entire book in a single day just once a week. A number of definitions include a reference to daily reading: an hour per evening (Krashen, 1981:105), a page a day (Matsumura, 1987:179) and 30 minutes per day (Dalle, 1988:25)

ER is uninstructed. Uninstructed reading (UR) refers to reading without assistance or the use of any tasks or explanations before or after reading. Extensive reading (ER) is uninstructed reading (UR) carried out extensively. UR may not necessarily be 'extensive' in that a student may read a single book without guidance or instruction (UR).

Reading with tasks (R+) is reading where a book may be discussed and pre- or post-reading tasks assigned. This is discussed further in Chapter 3.

1.3 The Power of Reading

Krashen (2004b) argues that evidence from numerous studies in first language (L1), second language (L2), and for children and adults shows that those who do more recreational reading (uninstructed reading) show better development in reading, writing, grammar and vocabulary. Smith (2006:12) suggests that the evidence is 'overwhelming'. Nagy and Herman (1987, quoted in Coady and Huckin, 1997:225) suggest that teachers

should promote extensive reading because it can lead to ‘greater vocabulary growth than any program of explicit instruction alone ever could’. ER is a pleasurable activity and the pleasure of reading can increase motivation and the desire to read more. With regular exposure to text at a comprehensible level (95% - 98%; Hu and Nation, 2000), readers are exposed again and again to language, and through this repeated exposure they are able to acquire the language in much the same way as they acquired their first language. ER has been shown to improve a variety of language skills in addition to developing general knowledge and improving attitudes to learning. Maley (2005: 354) concludes that extensive reading has been proclaimed as ‘the single most effective way to improve language proficiency’.

1.3.1 Gains from reading

With the reader exposed to a large quantity of text at a level where they comprehend approximately 95% - 98% of it without assistance (Liu and Nation, 1985; Laufer, 1989; Hu and Nation, 2000; Laufer and Ravenhorst-Kalovski, 2010), ER has been shown to lead to an increase in reading speed. In young adult subjects exposed to extensive reading, compared with slower intensive reading, Bell (2001) in Yemen, and Al Homoud and Schmitt (2009) in Saudi Arabia, demonstrated that the extensive reading groups achieved both significantly faster reading speeds and significantly higher scores

on measures of reading comprehension than the intensive reading groups. Lai (1993a) also found that extensive reading increased the reading speed of her secondary school students in H.K. Extensive reading exposes students to high frequency words repetitively, which can aid their word recognition, leading to automaticity (Jarrell, 2003; Nation, 2001). Automaticity leads to faster decoding of a text, which can aid comprehension (Bell, 2001).

Not only did Al Homoud and Schmitt (2009) note an improved reading speed for the ER students but also the extensive reading participants reported much more positive attitudes toward reading, their class, and their learning than the participants in the intensive reading group. Elley (1991:397) reported that the students involved in an ER programme developed very positive attitudes as their literacy levels from reading improved. Through the use of easily read, interesting materials, learners are able to experience success, which can motivate them to read (Waring, 2002). ER can also build confidence in the reading of longer texts, which is especially important for students such as those in H.K., who have mostly struggled through dense, intensive reading of short examination-type texts (Jarrell, 2003). Not only may ER improve attitudes towards language learning but McQuillan (1994) and Dupuy (1997) noted that students who participated in a course of grammar exercises or the extensive reading of popular

literature (perhaps unsurprisingly) favoured reading by a significant margin. A study by Elley and Mangubhai (1983) involving 380 students in eight rural Fijian schools demonstrated not only reading comprehension and listening gains for the experimental ER group but also a general improvement in Mathematics, General Studies, and even Fijian Language Intermediate Examinations, which the authors believed implied an incidental and positive change in attitude to school (pg. 65).

Elley (1991) reviews a number of studies with children between six and twelve years of age, in which subjects showed rapid growth in language development compared with learners in regular language programmes. There was a "spread of effect from reading competence to other language skills - writing, speaking and control over syntax," (Elley, 1991:404). Hafiz and Tudor (1989) in Pakistan, and Robb and Susser (1989) in Japan, also noted a writing improvement after an ER programme. Cho and Krashen (1994) reported that their four adult ESL learners increased competence in both listening and speaking abilities. Nation (1997) and Day and Bamford (1998) report that extensive reading resulted in students' making significant gains in other aspects of foreign language competence such as listening, writing, and vocabulary.

Grabe (2009) suggests that ER can lead to gains in world knowledge and conceptual knowledge and that these gains allow for success with more complex academic tasks and motivation for further learning, perhaps in part explaining the carry-over into other subject areas reported above by Elley and Mangubhai (1983). Yu (1993) found that subjects in a H.K. study of secondary students in a two-year extensive reading programme perceived that extensive reading lead to an increase in reading interest, vocabulary, reading and writing skills and world knowledge. This effect can be somewhat circular, as reading builds world knowledge (Pan, 1990) and Pang et al (2003) suggest that world knowledge may aid reading comprehension.

Extensive reading may provide a pleasurable way for students in poor L2 environments to obtain input. In fact, Nation (2001:155) argues that the use of reading and other input sources may be the only practical options for out-of-class language development for some learners, especially in EFL contexts such as that found in H.K.

ER is therefore able to help develop a variety of language skills, improve general world knowledge and aid motivation. Its potency lies in its being able to achieve these gains seemingly in the absence of direct instruction as uninstructed reading (UR). An

additional important attribute of ER is its ability to aid vocabulary acquisition, which is the focus of this study.

1.4 The introduction of ER in H.K.

In 1979, an assessment report on the standard of English in H.K. found that two-fifths of pupils from English medium schools (EMI) and four-fifths from Chinese-medium schools (CMI – Appendix B1) had not attained a standard of English that was acceptable to society, either educationally or for employment (Yu, 1979). There was a widely held perception among educators and employers that the English standard of students at all levels was falling (Education Commission, 1990). The Extensive Reading Scheme (HKERS) was first launched in 1991 in a handful of schools and implemented on a phased basis over seven years as a means to aid English language learning. An international study on reading literacy indicated that after H.K. students had learnt English for nine years, their mean scores were very low compared with international norms and far below the standard achieved by similar L2 students studying in Singapore (Johnson, 1994; Johnson and Cheung, 1995). In response, the HKERS was extended to primary schools in 1995.

In response to the perceived failure of students to acquire basic academic knowledge, particularly in the core subjects (science, maths, geography and history in addition to English and Chinese), the Hong Kong Government issued the Medium of Instruction Guidance for Secondary Schools in the early 1990s (Panel on Education, papers, 1997). The Guidance provided that all secondary schools should adopt Chinese to teach all academic subjects, starting with their Secondary 1 intake in the 1998/99 school year, and progressing each year to a higher level of secondary education, unless a school had obtained approval to use English as the medium of instruction (Appendix B1).

It was believed that by actively promoting Chinese as a medium of instruction (CMI), less importance would be attached to English. The government therefore allocated substantial additional resources to the teaching of English. One of these responses was the Hong Kong Extensive Reading scheme grant, introduced to provide funding for all primary and lower secondary schools to further develop and extend the H.K. extensive reading scheme. The scheme was based on the Edinburgh Project on Extensive Reading (EPER, 1998). In the 1998-99 academic year, 164 primary schools and 199 secondary schools participated in the scheme. Reading awards were offered to encourage students to read more (Legco Panel on Education, 1998; LEGCO – Legislative Council). The aim of the scheme was to boost students' English language capabilities, particularly

those of weaker students in CMI schools who would be exposed to less English due to the introduction of the new medium of instruction policy.

1.5 The failure of extensive reading in H.K.

A number of writers suggest that extensive reading in H.K. has not been particularly effective in improving students' language ability, particularly in the case of those weaker students in CMI schools who were the very ones the scheme was particularly designed to assist (Green, 2005; Wong, 2001).

Lai (1991, 1993a) carried out a one-year study involving 11- 13 year olds from eight H.K. secondary schools. Two low ability groups performed worse than the control group in general language proficiency. In reading comprehension, one low-level school performed significantly worse than the control school. Lai concludes that an increase in comprehensible input did not help students to improve their reading comprehension ability (1993a:29). In vocabulary recognition, a high-level group attained a lower score than the control group. The results do not consistently show gains for all groups.

Lai (1993b) reports on an extensive reading scheme that was piloted in 18 Hong Kong secondary schools in 1986 for a period of a year. Lai (1991:51) notes that only the more

able students from the experimental group progressed better in their reading competence than those in the control group. Lai notes that overall, the control groups performed better than the experimental students involved in ER.

The researcher's first-hand experience with ER in a H.K. school also suggested that the scheme was not very effective and highlighted the difficulties of implementing ER with lower ability students in a H.K. CMI school.

The researcher first entered a CMI Hong Kong secondary school in 2001, four years after the introduction of the extensive reading scheme grant and two or three years after the school had begun participating in the programme. The class had a box of books and comprehension and answers cards based on the Edinburgh Project on Extensive Reading (EPER) scheme. Few students in the class could read any of the books but often, somehow, managed to complete the comprehension worksheets, which seemed to be the main priority, usually by copying the answers from another student who had already 'corrected' their own work. Even after trying to add a large number of even easier books to the class box, most of the students still struggled to read the books unaided. No quantitative measure was carried out on students' acquisition of language from ER at this time.

Six reasons are suggested for the perceived failure of ER in Hong Kong, as outlined below. The first five could be classified as relating to a lack of belief or understanding as to the value of uninstructed reading over instructed language learning. The remaining reason could be classified as a linguistic or cognitive learning issue that may hinder acquisition from UR.

Firstly, ER requires that time is made available for it, which means that to be effective, it requires the support of the entire school community. Reading extensively for pleasure requires a certain change in the learning mind-set of Hong Kong students, teachers and parents. The students need to take responsibly for their own learning such that they self-select books and are motivated to take the time to read them without the need for constant assessment. Teachers need to believe that reading can improve students' language and thus provide them with materials, encouragement and time for the task. Parents need to understand that reading for pleasure may not be 'homework' in the traditional sense but can lead to long-term language gains.

In the researcher's experience, completing work from the various and numerous textbooks and workbooks on schedule was vital, and essentially the students' reading was to be completed at home. There was little actual time for ER, which was relegated

to the status of 'luxury' or 'optional extra'. Of 284 Hong Kong primary schools surveyed on ER (prior to the implementation of the primary scheme), only 4.7% said that students were given time during English lessons to read English books, but this was only done occasionally (Yu, 1993). In spite of the evidence in support of ER and its increasing popularity, Reynanda and Jacobs (2002) suggest that it is still often treated as a fringe issue in language programmes.

In reality, schools provided little time for ER. Green (2005:308) notes that schools interpret government-issued guidelines in different ways, with some marginalizing extensive reading to the extent that it occupied no space at all on the formal timetable but took on an extracurricular remedial status.

Secondly, the primary method of assessing students in H.K. is through examinations, which tend to measure discrete skills and require a great deal of specific teaching (Yu, 1979). Readers may not necessarily perform any better than non-readers in such tests and teachers may see little to be gained from extensive reading in the short term and often have little interest in the long term. Davis (1995:330) has noted that, 'the benefits [from extensive reading] do not emerge immediately in terms of straightforward examination results'. Cheung (1990) suggests that the majority of Hong Kong parents

are more concerned about their children getting ahead than in getting an education. A H.K. Institute of Education (HKIE, 1997) study of 4,172 primary schools found that in primary 1, 87.1% of students enjoyed reading for pleasure, but this had reduced to 18.9% by primary 6. The authors suggest (pg. 55) that the pressure of the examination-orientated system was to blame, with students having less time for reading, as they had important examinations to contend with.

Thirdly, extensive reading is student-centered, with students selecting what they want to read, and requires self-motivation to read (Grabe, 2000:326) and keep reading, with the end goal being to enjoy the task. Reading for pleasure largely lacks guided, controlled, follow-up tasks relying on implicit acquisition of language. In contrast, learning in Hong Kong is largely teacher-directed, with examinations driving motivation. The H.K. system has little time for pleasurable activities in a very exam-orientated implicit learning regime. Prowse (2002:144) sums up this problem when he notes that sometimes: ‘a class of students reading silently is not perceived as a class learning, let alone being taught, both by the students themselves and the school administration’.

Fourthly, students have little experience with ER in their own language and even less in English. Yu (2000:8) notes that of 537 young H.K. people interviewed (aged 10 - 24),

only 19% claimed that they read often (L.1) and 53% rarely or never (the figures may not be too dissimilar to other countries). In a study by Storey et al (1997), the percentage choosing English books ranged from 15.6% at secondary 1 to 4.6% at secondary 7 (year 13). A survey of the top 135 secondary one students (n = 210) at the researcher's CMI school at the start of the 2006 school year revealed that: 0.7% of students read often in English and 75% rarely or never. Wong (1993:57) notes that while English is easily accessible in H.K., it is neither extensively nor widely used by the majority of the local Chinese.

Fifthly, many students have very poor vocabulary knowledge and generally reading means translating every word into Chinese: thus, there is little comprehension or ability to read for meaning, and in fact, the students are largely reading in Chinese. Wong (2001) comments that the students rely on their dictionaries when they read, as they believe that knowing the meaning of every single word in their English reading will help their comprehension, which makes for a tedious and slow decoding process and ultimately a lack of pleasure in reading.

All of these factors may have in part contributed to the perceived failure of ER in H.K. and were certainly factors that the researcher noted as problems in his classroom. An

additional (sixth) factor that needs to be considered is that there may be some inability of H.K. Chinese students to acquire language simply through uninstructed reading. Is it possible that secondary school students may have experienced so many years of explicit teacher-directed learning that they may simply be incapable of learning any other way, or not desire to do so? Guthrie and Wigfield (2000) suggest that over time, some learners become less motivated to learn for the sake of learning and are driven by performance goals, which, in the case of pleasure reading, are usually intrinsic and therefore perhaps less relevant.

While studies involving students from many different countries have shown that they can acquire language from uninstructed reading, strategies that work in one educational, cultural, and linguistic context might not work in another (Gu, 2003).

A second point to be noted from the limited H.K. ER research is that not all students seemed capable of making language progress from ER. Lai (1993a) showed that the lower ability students performed poorly in general language proficiency and comprehension, although satisfactorily in vocabulary acquisition, where the higher ability students performed poorly. The H.K. pilot study (Lai, 1993b) suggested that only

more able students progressed in their reading competence, with few gains for less able students.

Dykes (2008a) notes that while research indicates that implicit language acquisition (the ability to acquire language through exposure to it) through uninstructed reading is possible, it is not clear if this is applicable for all Hong Kong school students and particularly for weaker students. Research has also suggested that a combination of implicit and explicit (language that is taught) learning, particularly of vocabulary, is more effective than either method in isolation (Nation, 2001; Nadarajan, 2009; Sonbul and Schmitt, 2009).

1.6 An alternative to uninstructed reading or ER

While teaching English in local schools in Brunei Darussalam, the researcher found that the textbook the students were required to use was generally too difficult and uninteresting for lower secondary classes. He started to use graded readers with worksheets. The whole class would read the same reader while being guided by the teacher and then complete activities designed to enhance vocabulary and basic language: filling in missing words, rearranging sentences, matching words with meanings, role plays, writing alternative endings, etc. It was found that the students enjoyed this

programme of guided reading with tasks (R+) and the programme was extended to include some upper secondary school classes with appropriate tasks.

UR in the form of ER requires a reader to read independently without teacher guidance or the use of tasks. Whereas ER is student-centered, requires self-motivation and has an end goal of pleasure, guided reading (R+) is teacher-centered and directed. In addition to the goal of pleasure reading, it also has an end goal of task completion. While the researcher had experienced the joy and benefits ER / UR himself as both a student and a teacher with first language students, his experience in H.K. was of a programme that had largely failed. The success of guided reading (R+) in Brunei caused the researcher to speculate whether R+ might also work in H.K. better than ER, particularly as it might fit the H.K. learning style better by being teacher-directed with clear learning outcomes through the completion of tasks.

Prior to implementing a full-scale ER programme in H.K., it needs to be determined whether a) H.K. secondary school students can acquire language (vocabulary) from uninstructed reading, b) if this method is suitable for students of all ability levels and c) if this reading method is the most effective for language acquisition or if an alternative method employing implicit and explicit learning such as R+ may be a better approach.

Vocabulary knowledge is a key indicator of language success and reading has been touted as one of the best ways of improving this knowledge (Bernhardt and Kamil, 1995; Laufer, 1992; Nation, 2001, 2006; Qian, 1999, 2002; Ulijn and Strother, 1990; Beck et al, 1982; Schmitt, 2008). In order to acquire vocabulary knowledge from extensive reading, it is necessary for the learner to be able to acquire at least some small vestige of vocabulary from a single reading. If some small measure of vocabulary acquisition can be detected from a single reading, one could conclude that this gain could be strengthened and multiplied through repeated reading, as in an extensive reading programme. If a study can demonstrate that a group of H.K. students of mixed ability can acquire vocabulary from UR, then it may be possible to conclude that previous failures of H.K. ER schemes were due to an implementation failure resulting in students' not engaging adequately with text.

Before putting time and money into an ER programme, it is vital to determine that the target population is able to acquire language from uninstructed learning. It is also vital to determine whether this style of learning is suitable for a range of ability levels and whether this particular method of 'learning' is the most effective.

To date, there has been little classroom research to compare uninstructed reading (UR) and reading with tasks (R+) and no research (to date) has considered whether UR or R+ is more suitable for a particular ability level of student.

This study therefore seeks answers to the following questions:

1. Can a group of weaker and stronger year 12 H.K. L2 Chinese students acquire vocabulary from uninstructed reading (UR) measured receptively after 14 days and 28 days?
2. Can the same two groups of students acquire vocabulary from reading that is supplemented with vocabulary tasks (R+) measured receptively after 14 days and 28 days?
3. From which of the two methods, UR or R+, can the subjects acquire the most vocabulary? In other words, which method is more effective for vocabulary acquisition?
4. Do students prefer R+ or UR and is there any difference in the opinions of WR and SR?
5. Is there a difference between the two methods (UR, R+) for weaker and stronger students to acquire vocabulary? In other words, can weaker students acquire

vocabulary from UR in the same quantity as stronger students and can weaker students acquire vocabulary from R+ in the same quantity as stronger students?

1.7 Study outline

Chapter 2 begins by outlining the importance of vocabulary for second language learners and their ability to acquire it from extensive reading. Problems with effectively administering an ER programme in H.K. are discussed and the failure of ER in H.K. is highlighted. The second section of chapter 2 examines a key issue in second language acquisition (SLA): whether the process of acquiring L2 is a subconscious (Krashen, 1989, 2004a) or a conscious (Schmidt, 1990) process. Several assumptions for ER / UR are explored in relation to subconscious acquisition. Krashen's Comprehension Hypothesis is discussed as a key theory in support of implicit learning and ER studies are reviewed. In response to arguments in sections 1 and 2, which suggest that vocabulary learning from UR has not been particularly successful in the H.K. context and may be haphazard, section 3 discusses reading with supplementation (R+) and the involvement load hypothesis (Laufer and Hulstijn, 2001). This section makes a case for R+ for vocabulary acquisition and concludes by reviewing studies for reading with tasks.

Chapter 3 is concerned with methodology and research design. The paradigm of the study is discussed, followed by a detailed description of the quasi-experiment, including participants, materials and research procedures. Chapter 3 concludes with a discussion on ethical and validity considerations. Chapter 4 analyses and explains data from the experiment in relation to the five research questions. Each research question is analysed under the headings of research analysis (RA) and the appropriate research question number. Chapter 5 provides a discussion of the results, contrasting them with other studies and the theoretical literature, and also examines the strengths and weakness of the experimental data obtained. Chapter 6 draws together the conclusions from chapter 5 and makes some recommendations for reading in H.K. and future research.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

Part One of this review begins by outlining the importance of vocabulary for second language learners and their ability to acquire it from reading. It is suggested that, at least mathematically, it may be possible for students to receive sufficient exposure to new words from uninstructed reading to be able to acquire them. It is argued, however, that UR is problematic in the H.K. context, where there is little incentive for students to read for pleasure. H.K. studies suggest that not all learners have been able to benefit from ER. There is a suggestion that reading supplemented with instruction or tasks may be more beneficial for vocabulary acquisition than UR alone.

A key assumption of UR is that language can be acquired implicitly. Part 2 examines the role of implicit and explicit learning, the role of noticing and four key assumptions for ER. While there are some studies that demonstrate modest vocabulary gains from UR, very few have explored whether these results are applicable to weaker readers and many such studies have involved university students. Krashen's Comprehension Hypothesis is examined as an important theory supporting UR and implicit acquisition

of language. Part Two concludes by examining studies in support of extensive reading, particularly those which highlight vocabulary acquisition.

Part Three compares and contrasts extensive reading and reading with tasks (R+). A rationale for R+ is provided and it is argued that it is an effective way to encourage students to employ an interactive reading model. The review argues that R+ is able to develop a reader's schema, which may be particularly important for weaker readers. The review concludes by discussing the role of tasks and the task load hypothesis before reviewing a number of studies employing R+ or contrasting R+ with other methods.

A number of the studies reviewed in this article involve university students. Tanaka and Stapleton (2007) note that while developments in foreign language reading in the Japanese context, for example, have witnessed considerable recent coverage, much of the research has been at the university level. Generalisations across age or ability groups are readily made in the L2 literature, particularly from predominantly quantitative research. Very few studies involving reading make use of random population samples and as quasi-experiments, all samples are in some way biased and any generalizations need to take this into account.

The age of a subject may affect the quality and quantity of language acquired from reading. Results from adult vocabulary acquisition studies may not be applicable to children or vice versa. DeKeyser (2000) suggests that age may have an effect on components of language that are learned implicitly, such as in UR. A contributing factor for this may be that children may have fewer inhibitions and learn more freely and naturally than adults, with less analysis or fewer inhibitions ('lower affective filter', to use Krashen's terminology - see 2.2.4). In short-term acquisition studies, older learners may out-perform younger learners, while the opposite may be true in longer-term studies.

Bialystok (1997) suggests that no maturational constraints are needed to account for differences between adult and child L2 acquisition but the difference may be due to processing differences. With a much larger L1 vocabulary, adults do not have to acquire new concepts in the L2, as children do, but only need to learn the verbal symbols representing these concepts.

Space constraints limit this argument but it should be noted that much of the evidence for vocabulary acquisition from UR has been obtained from studies with adults and university students, and thus may not be applicable to a younger population group, and

this needs to be considered particularly when making claims from one population group to another.

The quasi-experimental phase of this study involves senior secondary school students (M = 16.3 years old) and this review draws on literature from below and above this age group.

2.1.0 Vocabulary acquisition from uninstructed reading (Part1)

2.1.1 The importance of vocabulary

Extensive reading is able to help develop a variety of language skills, improve general world knowledge and aid motivation. Its potency lies in its being able achieve these gains seemingly in the absence of direct instruction as uninstructed reading. An additional important attribute of UR is its ability to aid vocabulary acquisition, which is the focus of this study.

Vocabulary is an essential part of mastering a second language (Schmitt 2008). Singleton (1999:4) states that ‘the major challenge of learning and using a language – whether as L1 or as L2 – lies not in the area of broad syntactic principles but in the

“nitty-gritty” of the lexicon’. Hunt and Beglar (2005:2) go as far as to argue that ‘the heart of language comprehension and use is the lexicon’, with Lewis (2000:8) suggesting that acquiring a sufficiently large vocabulary is the single most important task facing language learners, followed, according to Laufer and Sim (1985), by subject matter knowledge and syntactic structure. On the importance of vocabulary and grammar, Wilkins (1972:111) declares that, ‘....while without grammar very little can be conveyed, without vocabulary *nothing* can be conveyed’.

A number of writers (Nagy, 1988; Gu and Johnson, 1996; Grabe, 2009) note a positive correlation between reading proficiency / comprehension and vocabulary size. Qian (1999) found that scores on vocabulary size, depth of vocabulary knowledge and reading comprehension are highly and positively correlated. Stahl (1983) and others (Beck et al, 1982; Kameenui et al, 1982) note that an improvement in reading comprehension can be attributed to an increase in vocabulary knowledge. While many learners recognize the importance of vocabulary, it is at the same time one of the most difficult aspects of learning a language (Laufer, 1986). In the H.K. context, even after twelve years of language study, some students have not acquired the most common 1,000 English words, making further study difficult.

Some estimates (Nation, 2001; Grabe, 2009:269) suggest that the average (US) high school L1 graduate knows about 40,000 words and that L1 students gain a reading vocabulary of some 2,000 to 4,000 words per year. To learn to read 3,000 words a year would require about 70 words per week (44 weeks) and the default argument is that these words must have been largely acquired through uninstructed reading (Nagy et al, 1985; Coady and Nation, 1988). Nagy et al (1995) suggest that for L1 students, children's reading accounts for a substantial portion of the vocabulary growth that occurs during school years. Laufer (1992) suggests that knowing a minimum of about 3,000 words is required for effective reading at the university level, whereas knowing 5,000 words indicates likely academic success. Nation (2006) suggests that 8000 – 9000 word families are required to read authentic texts.

A rough estimate for EFL university students (Laufer, 2000) in Japan, China and Indonesia shows an average vocabulary size of 2380 words after about 1420 hours of instruction or about 7.5 years (at 4.5 hours a week / 44 week year), averaging about 7 words a week. The H.K. secondary school students involved in the current study had completed 11 years of schooling, which included, on average, at least four and a half hours of English a week, and yet 10% of them (n = 30) had not attained the 1000 most common English words (Nation, 2001), 53% had not attained the 2000 level and 70%

had not attained the 3000 word level with just a year and a half remaining before they hoped to enter university, leaving this particular H.K. sample well behind the necessary target. These sampled groups of L2 students have acquired vocabulary very slowly and even though they are at or about to start university, they have insufficient vocabulary to study in English at this level. Grabe (2009:273) suggests that students can only gain sufficient vocabulary for academic work through regular exposure to vocabulary through reading.

A number of writers (Nation, 2001; Stahl, 1999; Schmitt, 2000; Stahl and Nagy, 2006) suggest that if students read 1 million words and they know 95% – 98% of them, they would be exposed to 20,000 – 40,000 new words, and if they learned 1 in 10, they would acquire 2,000 – 4,000 new words. Using this calculation for the H.K. students involved in the current study reading a graded reader of about 13,000 words, they would be exposed to between 260 - 390 new words in each book and remember 26 – 39 of them. This would require them to read between 26 and 38 total books to acquire 1000 words, which could be possible in one to one-and-a-half school years (10 days to read a book). The reality, however, is that a survey of the H.K. subjects involved in the current study indicates that in the year prior to the study, the students had on average read about

2.4 books in English (n = 29), thus falling a long way short in both their vocabulary knowledge and in the quantity read (3.1.1).

The assumptions being made here are that vocabulary is crucial for language learners and although many L2 learners, including the H.K. students in this study, have not acquired sufficient vocabulary for language mastery, they can acquire it simply through a large quantity of uninstructed reading of appropriate texts. When about 98% of a text is understood, it is believed that readers are able to derive meaning of the 2% of unknown text from the known context.

2.1.2 Implementation problems of extensive reading

The idea of allowing students to read freely for pleasure in Hong Kong without producing an assessable product is quite foreign in a school system where everything is rigorously assessed. Lin (2001:85, 86) comments that public exam-taking skills constitute the most important factor for success in schools and society and the ‘rules of the game do not hinge on gaining an education or learning but on exam taking skills’. Kwan (1988; quoted in Chew, 2003:19) noted that teachers complained that students do not read unless there is a way of counting marks for the reading, and little had changed when Leung (2005:30) notes that final-year students (form 7 or year 13) felt that silent

reading in school was a waste of examination revision time. A Hong Kong study by Balla et al. (1991) showed that students had little incentive to undertake learning outside their immediate studies and tended to limit their work to what was specifically taught in a course. More recently, Braine (2009) reports on a survey of 30 Mainland students studying in H.K., of whom 7 were English majors, that most had never read for pleasure and had no use for English other than for academic purposes.

Because reading is not seen as a pleasurable leisure activity, there is little reading culture in either L1 or L2. The results of a survey released in April 2003, which covered 35 countries and 350,000 nine-year-olds, placed Hong Kong bottom of the league for primary students' interest in reading for pleasure in their own language (Hui, 2003). The author suggests that a lack of awareness of the importance of reading and the limited emphasis in schools may be partly to blame. It may be quite difficult to develop an interest in reading for pleasure in a second language without firstly developing it in one's first language. Day and Bamford (2000) note that teaching reading in a 'non-reading' culture, or in one that does not attach importance to reading for pleasure, makes the task of the EFL reading teacher more complex.

The lack of a L2 reading culture is evident in a study by Yang (2001) of 120 adults enrolled in university evening classes in Hong Kong, which found that most of them had never read an English novel in spite of the fact that they had all graduated from secondary school and some had had a tertiary education experience. Yang (2007) suggests that students rarely read for pleasure, as they think that reading is an aspect of textbook-related activity. Only one out of five children in Hong Kong said they read outside school for pleasure (L1) even after several years of exposure to the H.K. extensive reading scheme (Hui, 2003). It was suggested that lack of parental awareness of the importance of encouraging their children to read and the limited emphasis placed on reading at school were among the key factors accounting for the phenomenon and the poor performance noted above in an international study on interest in reading.

Pierson (1988) highlights a paradox in the H.K. educational process where after an investment of almost a decade and a half of study, students obtain entry to a tertiary institute in which English is the major medium of instruction and yet, even though they are talented, disciplined and conscientious in their studies, they simply do not read English books for leisure, pleasure or interest.

In addition to a lack of reading culture, Hamp-Lyons (1983:304) found that negative transfer from L1 reading affected the way students tackled L2 ER. The Malaysian students in the study had been taught to understand every word in an L1 text and they applied this method when reading in L2. Wong (2001) noted a similar problem in H.K. with students wanting to know the meaning of every single word they read, leading to very slow decoding, which hindered global comprehension and ultimately the ability to gain pleasure from a text. The researcher's students reported that their teacher of Chinese language had told them to find the most difficult reading book in Chinese, as this would help their Chinese better than an easier book, and the students transferred this technique to their L2 reading

Where extensive reading largely relies on acquiring language implicitly (Krashen, 1989, 2004a), the Hong Kong system is heavily weighted towards students obtaining language through learning explicitly through controlled, structured teaching (Green, 2005; Lai, 1991, 1993a; Wong, 2001). ER has therefore been largely seen as an isolated programme without clear learning goals for students (Green, 1995). Green suggests that schools that do allow time for reading in class do so without any form of conferencing, discussion or motivating of students to read, such that the sessions become like 'monastic detention sessions' (pg. 308). In the researcher's school, reading

comprehension activities and/or book reports were required after reading in order to have something to assess and measure to make the reading 'purposeful'. Students who struggle to actually read a book in L2 may have even greater difficulties in the follow-up task of writing the report that inevitably follows, raising the affective filter (increasing apprehension: see 2.2.4).

Wigfield et al (1997) and McKenna (2001) suggest that for L1 elementary level students, positive motivation for reading declines through the grades, a phenomenon also noted in L2 learners around the world (Dornyei, 2005). Yu (2000:255), in a study of more than 3,000 H.K. secondary school students, found that those in the first year demonstrated more positive reading attitudes and were more inclined to adopt reading as a favourite pastime than were those in the second and third years.

While motivation through the secondary-school grades does not decline consistently, there are no points at which it increases (Grabe, 2009). Guthrie and Wigfield (2000) suggest that as intrinsic motivation decreases, extrinsic motivation may increase and the desire to learn decreases while the desire for external recognition increases. In other words, learners become less motivated to learn for the sake of learning and are driven by performance goals. This suggests that an ER programme which relies on reading

solely for pleasure rather than reading and completing tasks could be more difficult to implement in a H.K. secondary school. While ER can help L1 learners to acquire vocabulary and has been shown to be effective for many L2 learners, it is somewhat problematic for H.K. L2 learners. To acquire language from reading, students need to be sufficiently motivated to interact regularly with books.

The mere introduction of a large number of high interest books in itself may not necessarily guarantee that students' language will improve. In Ethiopia, Ambatchew (2004) compared two experimental primary schools that had received a large supply of high-interest books with two schools that had not (n = 454). The programme evaluation concludes that the introduction of the books had no significant effect on the reading skills of the students. For a number of reasons, including a lack of motivation to read, the students failed to interact with the books and therefore failed to make gains.

In Malawi, book boxes were delivered to every year 4 and 5 primary class in the country (Williams 2007). End-of-project test results indicate a statistically significant decrease in mean scores on an informal reading comprehension test and a cloze for the students involved. It is suggested that not only were there deficiencies in programme

implementation but also cultural-educational difficulties, which inhibited students from effective interaction with the books provided.

2.1.3 Failure to acquire language from uninstructed reading

One of the problems with quantitative research, which has been predominant in L2 reading studies, is the tendency towards generalizations from statistical averages. Carter and Hurtado (2007:30) note that quantitative researchers tend to view subjects outside of 'the average' as anomalies and that these departures from average norms suggest habits or experiences of students that cannot conform to a 'one-size-fits-all' model. However, these anomalies are often not reported in quantitative studies but rather the norms or average gains of students are reported. While Elley and Mangubhai (1983:65), in a study of several hundred L2 readers in Fiji, showed significant language gains for ER, there are still large numbers of negative scores that largely go uninvestigated and unexplained. The generalized conclusion stands, however, that *all* students can improve their language through exposure to high-interest books. A large-scale study in Sri Lanka (Kuruppu, 2001) reported positive gains for students involved in reading compared with a non-reading control group. While there was a positive effect for reading overall, the results suggest zero or negative scores for some students. The Colombo grade 4 project,

for example, recorded a reading mean test gain of 10.51% and standard deviation of 9.04.

The failure of ER to meet the needs of H.K. students at all ability levels is highlighted in a pilot study of the H.K. extensive reading scheme carried out in 1986 (Lai, 1993b). Eighteen schools were matched in curriculum type (Anglo-Chinese or Chinese Middle) and in English proficiency level (high, mid, low) to form nine experimental groups (n = 1454) and nine control groups (n = 1266). Generally the students in the Anglo-Chinese schools, where there is an emphasis on English as a medium of instruction (EMI) for all subjects, are more proficient than those in Chinese middle schools, where students study all their subjects in Chinese (CMI - Appendix B1). ER was implemented in the experimental schools for a year, with three lessons a week taken from the usual English language teaching schedule. The scheme made use of graded readers with comprehension questions developed by the Institute of Applied Linguistic Studies of the University of Edinburgh. Both the control and experimental schools had pre- and post-tests on English proficiency and reading comprehension using the Hong Kong Attainment Tests and the Standardized Reading Test for Hong Kong (Irvine, 2006).

Results showed that, as a whole, the experimental and control groups had progressed at the same rate in English proficiency and reading comprehension over the year. However, when subjects were divided into two curriculum types (CMI and EMI) and three ability groups, it was found that experimental subjects in Chinese Middle schools (CMI) and those in the high English ability group had made a greater improvement in reading comprehension than the control (significant at 0.05). An explanation for this is that Chinese Middle school (CMI) students had relatively limited exposure to English compared to Anglo-Chinese school students and therefore the effect of extensive reading was more prominent (pg. 88). The reading programme was able to benefit the students more than the regular timetabled classes they had missed. As for the high English ability group, Lai suggests that they were possibly more motivated and hence could benefit from the scheme (Lai, 1991:51).

In general English proficiency, the experimental group made no gains over the control. In fact, they did worse than the control in language structures, reading comprehension and listening comprehension (all significant at 0.05). They attained the same average as the control in sentence writing. Lai (1991:51) notes that only the more able students from the experimental group progressed better in their reading competence than those in the control group and only students in Chinese Middle schools made significant gains in

reading proficiency. Lai notes that overall, the control group performed better than the experimental students involved in ER.

On average, students only read 19.6 books over the course of the year's trial ($SD = 10.4$), which Lai suggests may not have been enough to have made a difference, particularly as the experimental class had three English lessons taken away for extensive reading. The ER programme may have provided useful input for the students who read a lot but very little input for those with little engagement, particularly as these students had lost three normal class lessons. The control group had the three English lessons under teacher instruction, which may have resulted in these students receiving as much language input as the experimental group. The control group could also have made use of ER in their own time if they wished and the study did not attempt to control or measure this. Lai suggests that it is not firmly established that the more reading one does the better one acquires the language or if exposure to text is the ideal condition for language acquisition (pg. 52). In other words, the relative effects of the quantity and quality of reading input have not been established. It could be argued that while quantity of input is important, quality of input processed may be more important (Koda, 2007). Wang and Guthrie (2004) found that with fourth graders in Taiwan and the U.S., intrinsic

motivation strongly predicted text comprehension and also predicted the amount of reading done by students.

In Lai's (1993b) summer holiday Hong Kong study, students paid to attend the course in the summer holidays. 21% of the students said that they joined the course out of their own initiative, and 77% said that they joined as a result of parents' recommendation and their own initiative, possibly indicating a degree of intrinsic motivation. The students came from a large number of different schools from secondary 1 – 3 (S1 - S3, grades 7 - 9) with an average age of 13. There were 126 students in 1988 (S1), 88 in 1989 (S2) and 52 in 1991 (S3). Because of the special nature of this course, no control groups could be identified; however, the results of these students were compared with those of the control subjects from the H.K. ER trial study mentioned earlier (Lai, 1993b).

Students were placed into classes of about 20 students of similar reading ability, determined by a placement test, in order to provide them with a selection of readers at a suitable level. The classes ran for 4 weeks from 9 – 11.30 am. Apart from extensive reading, which was carried out at the students' own pace with books of their own choice, there were whole-class activities for half of the lesson: reading short passages or funny stories, singing songs, reading poems aloud, playing language games and working on

riddles or puzzles. The programme was thus not one of pure ER or UR. On average, the students in S1 read 16 books, S2 – 18.5 and S3 - 14.2 books. A Standardized Reading Test developed by the Hong Kong Reading Association was used for assessing reading comprehension. T-Test results of pre- and post-test comprehension scores of S1 and S2 showed significant gains (for S1, $t = 5.26$, $df = 125$, $p = 0.000$, and for S2, $t = 9.55$, $df = 87$, $p = 0.000$). Lai suggests that a lack of motivation by the S3 students resulted in less reading and no significant improvement in reading comprehension. The S3 students had lower average pre-test scores, 40.0, as compared to S1 – 44.2 and S2 - 47.3. It is difficult to draw conclusions from this study about ER, as the study combined ER with activities, except to say that less reading resulted in a lower final reading comprehension score. The weakest students (S3) were less motivated and read the least and also failed to make any significant language improvements.

While these H.K. studies do not specifically measure vocabulary gains from reading, they do measure reading comprehension or general language proficiency gains, or lack of, attributable to reading. There is a suggestion that any gains may be partially attributable to changes in vocabulary knowledge due to the correlation between reading comprehension / proficiency and vocabulary knowledge (2.1.1). One of the difficulties in measuring general language proficiency in longitudinal studies is the opportunity for

subjects to have obtained language from sources other than from reading. An additional problem is whether the measuring device (H.K. Attainment / Standardized Test) is sufficiently sensitive to detect changes occurring over the somewhat limited period of the study. The advantages for the current study are that the acquisition of specific vocabulary words are measured within a group on a pre-test / post-test capable of detecting small differences. The sensitivity of Lai's recording device is possibly strengthened by her relatively large sample size compared to the chances of a Type II error occurring with the small sample size in the current study (failure to detect a difference).

The H.K. studies suggest that if language is to be acquired from uninstructed reading, then a sufficient quantity of reading needs to be completed, which requires students to be sufficiently motivated; however, motivation for pleasure reading is not the norm in H.K. The second point is that language acquisition from reading does not seem to be guaranteed from ER and that weaker students in particular may not benefit.

In many instances, reading programmes have been introduced into L2 curricula to meet a perceived need to counter falling English standards or to raise English levels more effectively than through purely traditional classes (Fiji – Elley and Mangubhai, 1983;

Hong Kong – Lai, 1991; Singapore – Ng and Sullivan, 2001; Ethiopia – Ambatchew, 2003; Brunei – Ng, 2001; Malawi – Williams, 2007).

Elley states that

increased literacy learning through exposure to high interest books can occur under a range of circumstances regardless of the student's first language, age, culture of the classroom, number or types of books employed, training methods of the teachers involved or the research design or evaluation methods employed.

(Elley, 2001:238)

Elley states that increased literacy 'can' occur, not that it 'will' or 'must', and in his list of beneficiaries, he does not specifically mention students of 'any ability level', although perhaps that is implied under age level.

If ER is to be an effective tool in aiding L2 language acquisition in a nationwide programme, or even a single class programme, then it needs to be of benefit to all learners. Susser and Robb (1990:8) suggest that students with a 'certain level of ability' in English can learn to read by extensive reading alone, which suggests that they are able to acquire vocabulary. It is unclear, however, what this 'certain ability level' is, but

Nuttall (1996:127) describes many weak L2 students as being caught up in the 'vicious circle of the weak reader.' This means that they lack the vocabulary to 'get started' and therefore cannot comprehend the text and do not acquire any language. Coady (1997:229) suggested that there is a threshold level of vocabulary knowledge below which a learner cannot read well enough to learn new vocabulary through reading, which he termed the 'beginner's paradox' Both Liu and Nation (1985) and Laufer (1997) maintained that learners who have not yet reached a minimum of 3,000 word families (affording 95% coverage of most general texts) cannot adequately guess meaning from context in unsimplified texts, and must therefore learn many of these words through direct instruction. However, there is no set word level suggested for the reading of the simplified graded readers that are readily available nowadays for L2 learners.

Horst et al. (1998) found that in a study of 34 low intermediate level university students in Oman, students acquired 22% (1 in 5) of the target words. The study made use of a text that was read to students while they followed along. The text (*The Mayor of Casterbridge*) contained 21,232 words with 45 target words. The students' initial word knowledge as measured by a Levels Test (Nation, 1990) and the considerable variance in post-test scores suggest that some students in the group must have found the book challenging, which suggests that it was not at an ideal level: 98% comprehensible. Thus,

while students did make vocabulary gains, Horst et al. suggest that incidental vocabulary acquisition was not sufficient for the low-level learners to build up their lexicons (pg. 220). The ‘low-level’ learners in this study were recognized as those with a mean vocabulary size of around 3,000 words, much higher than for many students in H.K. CMI schools (for a comparison with the current study, see 3.1.1). This study also measures students’ initial word knowledge on Nation’s Levels Test (Appendix A1) but then matches the students with a text at an appropriate level. Horst et al used the same text for all their subjects even though the levels test had shown a wide range of abilities.

Mason and Krashen (1997 – experiment 1) conducted one of the few ER studies to have specifically explored the relationship between a subject’s ability level and their ability to acquire language from uninstructed reading. The 30 students were second, third and fourth-year Japanese students who had failed EFL classes and were placed into a ‘retakers’ (*Sai Rishu*) class. The control group consisted of second-year students in the general education curriculum. For the first semester, both groups followed the same 90-minute traditional curriculum, which included reading selections, vocabulary and grammar exercises, translation exercises and comprehension questions. Students in the control group turned in assignments on time, had almost perfect attendance and did well

in exams. The *Sai Rishu* group mostly failed to turn in complete homework and attendance was poor, with a third of the students dropping out before the end of the semester.

While the programme continued for the control group, the experimental group, *Sai Rishu*, spent the second semester reading graded readers. On average they read (*about*) 30 books, (no *SD* provided), wrote short synopses and recorded their feelings and opinions about their reading in a diary in Japanese. By means of comparison, in Lai's H.K. study (1991) students read on average 19.6 books (*SD* = 10.4). Mason and Krashen do not provide an explanation as to how they managed to get this group of failing students to read so readily.

A 100-item cloze test was given to both groups as a pre-post test. The test made use of a 1600-word passage with words deleted. It would appear that the same passage was used for the pre- and post-tests. Even if different words were deleted, the students had had greater exposure to the content for the post-test, which may have aided comprehension. The authors report that twenty subjects were randomly selected from each group for the study, but do not explain why this number was decided upon. The gains made by the experimental group were significantly greater than the gains made by the comparison

group ($t = 2.269$, $df = 38$, $p < 0.025$) and in fact the experimental group almost made up the initial gap between the two groups. The study does not report what type of t-test was used. While Mason and Krashen do not provide an effect size, it is calculated as $r = .170$, which is a medium effect. The comparison (control) group who continued with their regular classroom programme made a mean gain from pre- to post-test of 4.35 words without the benefits of the reading programme, suggesting that perhaps the measurement tool was not particularly accurate. The authors do not provide sufficient data to determine whether in fact the comparison group were also able to make a significant vocabulary gain without being involved in the treatment programme.

The authors report that many of the students became 'eager readers' by the end of the semester (pg. 93). It may be argued that gains with weaker students are to be expected compared to gains for stronger students. From a research perspective, this may be a valid argument; however, from a pedagogical perspective, ER provided these students with something that several years of previous classes had failed to do: success. While the study demonstrates gains for these 'weak' students, their actual ability level is not quantified. They were significantly weaker than the control group on a cloze test and had not done well on their university examinations, but we do not know, for example, the extent of their English vocabulary knowledge, so it is difficult to generalize from

this sample to other populations of ‘weak’ students. The study shows that the students were capable of greater language gains from UR compared with a traditional language programme, but alternatively, a language rich programme could demonstrate even greater gains.

A study by Zahar et al (2001) and a replication by Tekmen and Daloğlu (2006) both explored vocabulary acquisition from a single text (*The Golden Fleece*) for a mixed ability group of students. While the studies compared vocabulary acquisition between weaker and stronger students, the weakest groups are only relatively weak when compared to each other. Zahar et al employed 144 grade 7 ESL Francophone students. Of the weakest group, 56% showed mastery at the 3,000 Nation (1990) word level and 40% at the 5,000 word level, meaning that this group were quite advanced ESL students. Tekmen and Daloğlu employed university intensive English students, of whom 51% had obtained the 2,000-word level (level 2) and 41% at the 3,000-word level on the Nation (2001) word test. (By means of comparison, 17% of the stronger students in the current H.K. study and none of the weaker readers had attained level 2 after 11 years of English instruction: see table 3.4).

Zahar et al recorded the smallest percentage gains for the strongest and weakest ability students (although there was a possible ceiling effect for the stronger group) while Tekmen and Daloğlu showed the greatest gains for the stronger group (5.32 words or 17.7%), although gains were still made for the weaker group (10.4%). While the weaker students in the Zahar et al (2001) study could acquire vocabulary from the text, they needed more encounters to learn words than the more proficient learners. This seems to fit the maxim that the more you know, the easier it is to learn.

While the studies are valid for uninstructed reading, the subjects were not entirely unguided, as the reading involved listening while following the story, which guaranteed exposure to the text and removed the need for students to be self-motivated. Students were then able to read the story again on their own: thus, they were exposed twice to the same text (repeated reading). While the studies attempt to draw conclusions between vocabulary acquisition and learner ability, the text was more difficult for some learners than it was for others. For students for whom the text was at an optimum 98% comprehensibility, there may have been greater opportunities for acquisition than for students with a comprehension level of less than 95%, making group comparisons unreliable. Both weaker and stronger students were, however, able to acquire some

vocabulary, as measured on an immediate post-test. If the text were at a more appropriate level, the weaker readers may have made greater gains.

Zahar et al (2001) suggest that because the rate of acquiring vocabulary implicitly from UR is low, some form of explicit instruction is needed to get many students over the 3000-word threshold where they have some chance of reading texts independently and better acquiring vocabulary on their own. They suggest that reading be supplemented by either direct vocabulary instruction (Nation and Waring, 1997) or instructionally enhanced reading (Hulstijn et al, 1996). This thesis compares reading without supplementation or instruction, with reading which employs vocabulary tasks.

2.1.4 Summary of Part One

This section has outlined the importance of vocabulary and the means by which students may obtain the necessary vocabulary from extensive reading. It was suggested that UR in the form of ER is not generally seen as worthwhile, useful or particularly relevant in the H.K. context and H.K. students do not readily read for pleasure. There is a suggestion that weaker students in H.K. have not benefited from ER, but it is not clear if this is due to their lack of motivation to read and exposure to text or their lack of ability to acquire language implicitly from UR. It is suggested that the style of learning

required for ER, particularly that of being uninstructed, does not easily fit the traditional H.K. learning culture, which is not particularly student-centered and relies primarily on direct instruction for language learning rather than implicit acquisition. Few studies have deliberately sought to determine whether WR can acquire language from UR. The section concluded with a suggestion that supplemented reading may be more effective than UR for vocabulary acquisition, particularly for weaker readers.

A major premise for the effectiveness of UR is that learners can acquire language implicitly or subconsciously. The following section compares and contrasts implicit and explicit acquisition and discusses four key assumptions of ER and Krashen's Comprehension Hypothesis. The section concludes with a review of studies involving UR.

2.2.0 Language theories of uninstructed reading (Part Two)

2.2.1 Implicit and explicit learning

A key issue in second language acquisition (SLA) is whether the process of acquiring L2 is a subconscious (Krashen, 1989, 2004a) or a conscious (Schmidt, 1990) process. The claim for uninstructed reading is that language can be acquired subconsciously.

Gass and Selinker (2001) define implicit learning as that which takes place naturally, simply and without conscious operations or (DeKeyser, 2003) learning without awareness. Explicit learning is a more conscious operation where learners search for structure by making and testing hypotheses to gain meaning from language, be it the syntax or lexicon.

Doughty and Williams (1998; quoted in Hunt and Beglar, 2005:2) contrast the two by describing the goal of explicit teaching as being to ‘direct learner attention’, whereas the aim of an implicit focus on form is to ‘attract learner attention’ while ‘minimizing any interruption to the communication of meaning’.

Laufer and Hulstijn (2001:11) note that implicit learning can only be incidental without learners’ deliberate decision to commit information to memory. Incidental learning can often be viewed as a secondary result of the primary goal of communication (Hunt and Beglar, 2005). Laufer and Hulstijn (2001) note that explicit learning can take place both intentionally and incidentally, where intentional learning is a deliberate attempt to commit new information to memory. A vocabulary task that requires linking word form to word meaning is an explicit learning activity, requiring attention on the part of the learner, but the vocabulary can be learnt either intentionally or incidentally.

There are three positions that are commonly distinguished to explain the question of how second language is acquired: the Strong Interface, Non Interface and Weak Interface positions.

The Strong Interface position suggests that explicit knowledge may transfer into implicit knowledge over time (Sharwood-Smith, 1993, quoted in Laufer and Hulstijn, 2001). DeKeyser (1997) contends that the practice of learned language can lead to automaticity over time as practice bridges the gap between explicit knowledge and use, or, as DeKeyser notes (2003), explicit learning can become implicit as learners lose awareness of the rule but retain the use of the structure. (Automaticity is fluency, where the learner can convert language into meaning without conscious thought). In this position, providing students with an opportunity for practice or output is important.

The Non Interface position contends that while explicit metalinguistic knowledge (knowledge of rules) does not affect the acquisition of implicit knowledge (Krashen, 1989), it cannot become implicit knowledge.

Krashen (1981) claims that knowledge of consciously learned language is stored differently from unconsciously acquired knowledge and that only the latter is employed

in spontaneous language use and there is no interaction between these two systems. This means that learned language can never be used spontaneously, as acquired language can. Krashen's recommendation, therefore, is that L2 students should have as much access to comprehensible input as possible for implicit learning and do not require explicit rules and systematic practice of these rules. He argues that free voluntary reading (ER) is one of the best means by which to obtain comprehensible input. Output is therefore not important for acquisition.

Swain and Lapkin (1995), however, contend that explicit learning is useful for at least some rules, and Swain (1996) suggests that it is the role of practice to close the gap between explicit knowledge and use.

The Weak Interface position contends that explicit learning may indirectly affect the acquisition of implicit knowledge by focusing learners' attention on features of input that are critical for language to be acquired (Ellis, 2001). The Weak Interface position was further developed into the Noticing Hypothesis by Schmidt (1990).

2.2.2 The noticing hypothesis and key assumptions of uninstructed reading

The Noticing Hypothesis contends that for L2 grammar acquisition to take place, it is not necessary to learn all the rules, but that learners must 'notice' critical features or mentally register an event. Schmidt contends that many features in L2 input may be communicatively redundant or infrequent and that learners who take a passive approach to learning by waiting patiently for something to trigger automatic noticing are likely to be slow and unsuccessful learners. Schmidt argues that there is no learning without attention, which is necessary for explicit learning and may be both necessary and sufficient for implicit learning. If a learner is not consciously aware of a specific language feature - i.e., is unable to articulate that it is problematic - then he or she will not be able to learn that language feature, whether grammatical, lexical or pragmatic (Schmidt, 1990:129).

Laufer (2003) notes that a learner being able to 'notice' an unknown word is the first of four assumptions required for implicit acquisition of language from reading. The assumptions are that a reader must a) notice an unknown word, b) correctly guess its meaning, c) retain the correctly guessed word and d) that word acquisition is cumulative.

An essential condition for acquiring a word is that a learner must notice or pay attention to new material (Schmidt 1994). Paradis (2000; quoted in Laufer 2003:16) notes that those who argue that learning can occur without attention do so for grammar, not vocabulary, where it is essential. A key element of noticing is that a learner realizes that an unknown word is in fact unknown.

A study by Laufer and Yano (2001), however, showed that learners over-estimated the number of words they knew by over 60%. One reason suggested for this is that the learners are unaware that they do not know a word, as they may confuse it with a word they do know, for example: embrace and embarrass, evaluate/evacuate, intimate/intimidate or privacy/piracy. Folse (2004) suggests that partial or completely incorrect inferences may be made, which would be especially unhelpful, particularly for weak learners. Failure to infer correctly may be contributed to four factors, according to Koda (2005), all of which are particularly common in weaker readers: poor vocabulary knowledge, insufficient background knowledge, inefficient word recognition skills and syntactic parsing skills.

The second assumption required for implicit acquisition of language from reading is that for noticing to lead to word acquisition, the learner must be able to guess the

meaning of unknown words after noticing them. The theory suggests that where most of the text is comprehended, the reader is simply able to guess the meanings of the words they do not know. It is possible, however, for the context to be of no help when guessing a word. Bensoussan and Laufer (1984) found that in a text for adult ESL learners, readers often had no contextual cues to the meaning of the unknown words and even when cues were available they did not always help, as wrong guesses occurred more frequently than correct and approximate guesses. Liu and Nation (1985), in a study of 59 English language teachers, found that while higher proficiency learners successfully guessed 85% – 100% of the unknown words, the lowest proficiency learners only guessed 30% – 40% successfully. Reading for general meaning, therefore, as in UR or ER, may not automatically lead to a reader correctly guessing the meaning of a word and their acquiring it (Huckin and Coady, 1999:182 and 189) and acquisition from UR may be more difficult for weaker readers.

If the reader has noticed an unknown word and determined that the word is important, then they may have attempted to guess its meaning (Hulstijn et al, 1996). Some of the words may be guessed correctly due to contextual aids or background knowledge of the reader; however, words that are easily guessed may not be retained as well as words that are guessed with greater difficulty (Mondria and Wit de Boer, 1991). Haastруп (1991)

argues that words that are more difficult to guess require greater processing effort and the more distinctive memory traces created by them leads to their being retained better in memory. It may not be so likely that readers taking part in 'pleasure reading' (ER) will invest time in trying to guess the meanings of 'difficult' words.

If the context is rich enough, then learners may simply miss out unknown words rather than needing to guess them, as their meaning may not be necessary for general understanding of the text (Mondria and Wit-de Boer, 1991; Fraser, 1999; Paribakht and Wesche, 1999; Zahar et al, 2001). This technique of skipping unknown words in favour of meaning to keep the reading interesting and enjoyable has been taught in teacher training courses and promoted in some ER programmes, perhaps in part from Goodman's (1968) view of reading as a psycholinguistic guessing game. Day and Bamford (1998:93) note that 'part of fluent and effective reading involves the reader ignoring unknown words and phrases, or if understanding them is essential, guessing their approximate meanings'. Birch (2007:148), however, suggests that the technique of skipping unknown words could result in the reader never acquiring the skipped words. If, however, readers always skip unknown words and never acquire them, then the reader must never progress to a more advanced level while extensively reading for pleasure in opposition to the comprehensible input theory of $i+1$, which suggests that

input at a level just slightly above the reader's current level will result in acquisition of new material (Krashen 1989, 2003 - see 2.2.4). It is suggested, therefore, that while some readers may skip some unknown words some of the time, other unknown words must be noticed and the reader must make an attempt at a meaning, as a number of studies point to some successful acquisition from UR. However, noticing and correctly inferring more difficult word meanings may be difficult for weaker readers. While H.K. students may look up each and every unknown word in a dictionary (Wong, 2001), this can reduce the total quantity of text students are exposed to by reducing their reading speed and limiting their enjoyment of books and thus their desire to read.

The third assumption for implicit acquisition of language from reading is that after noticing and correctly guessing the meaning of a word, the reader needs to be able to retain the correctly guessed word and meaning. Retention, measured over time, enables knowledge of words to be strengthened in subsequent meetings. A number of writers suggest that the actual number of new words retained from a single reading is about 5 – 15% (Nagy, 1997; Nation, 2001; Stahl and Nagy, 2006; Swanborn and de Glopper, 1999).

While a number of studies have measured general language gains from reading such as reading comprehension, writing or language proficiency gains, there are relatively few studies that have measured the acquisition of vocabulary from reading. Horst (2005) records only nine such studies, two of which are for languages other than English.

Waring and Takaki (2003:132) report on six 'oft cited' studies of vocabulary growth from reading in a foreign language (Day et al, 1991; Pitts et al, 1989 – two experiments; Horst et al, 1998). Two of the studies (Dupuy and Krashen, 1993 – French; Hulstijn, 1992 – Dutch) were for languages other than English and are not included in this review.

Day et al (1991) showed a very small short-term vocabulary gain for Japanese high school students doing ER but did not make a comparison with other methods. In this study, only the reading experimental group was exposed to the targeted tested vocabulary of words a pilot study had shown that the students did not know, by reading an article, and there was no pre-test to determine group differences. The post-test was multiple choice (MC) and only included 17 items. Multiple choice (MC) tests, particularly those employing such a small sample, are not particularly accurate for assessing how much vocabulary has been learned (Meara and Buxton, 1987; Wesche

and Paribakht, 1996; Waring, 1999). The significant gains made by the control group mean that we can conclude that some vocabulary acquisition is possible from ER, but it does not tell us if it is better than any other method. This thesis employs 26 target words with 15 distracters on a word-meaning matching test (Appendix A7).

Pitts et al (1989), in two experiments, tested ESL students ($n = 16$ and $n = 35$) on 30 nadsat (made-up) words from two chapters of *A Clockwork Orange*. The study did show small vocabulary gains, although it also used an MC test and employed a very difficult text, which 50% of the pupils failed to finish. The study reports high SDs, and low means which suggest that many students had zero scores ($M = 1.81$ $SD = 4.26$; $M = 2.42$, $SD = 2.62$).

The study by Horst et al (1998, see 2.1.3) found that out of 23 new words available for learning, 5 words were learned (22%) as measured on an immediate post-test. While an immediate post-test provides an indication of vocabulary acquired, it may be of little practical help without retention information. A word recorded as ‘acquired’ may be only loosely stored in short-term memory and it is repetition of the word through subsequent meetings that strengthens retention (Waring and Takaki, 2003; Joe, 2010). Horst et al may state that 22% of words were learned, but if this learning only lasts a day or two,

then there is little chance for the new words to be strengthened through additional encounters and possibly little practical gain. This thesis tests vocabulary after 14 and 28 days.

Waring and Takaki (2003), in a study of Japanese university students ($n = 15$), reported small gains on an immediate post-test of substitute words (nadsat words) from a single text (*The Little Princess*). More frequent words were more likely to be learned and were more resistant to decay, but on average, the meaning of only one of the 25 items was remembered after three months, and the meanings of none of the items that were met fewer than eight times were remembered three months later. On average, based on the three-month test, the students learned one word for each hour of reading. While this study suggests that very few words are acquired incidentally from UR, in the form of ER there may be opportunities for some of the newly acquired words to be encountered again within three months, thus strengthening knowledge of the word. This thesis employs 'real words' which means that there is a risk of the words being encountered outside the study, however there are ethical issues in using nadsat words and whether we want learners to spend time learning them.

Brown et al (2008), in a study of Japanese university students (n = 35) on UR, reported a 35% (4.10 words from 28) gain on an immediate prompted (MC) post-test with little decay after one week (-.08) and three months (-1.17). An unprompted (translation) test, however, showed a gain of only 3.5% (one word) after three months. This is perhaps one of the most comprehensive experiments measuring vocabulary gains from reading. The variation in scores between the two test types (MC and translation) underlines the importance of multiple testing in making claims for acquisition. This thesis, however, only employs a single test as it was felt that multiple testing would extend the test time such that fatigue could affect the scores.

Swanborn and de Glopper (1999), in a meta-analysis of 20 studies involving native speakers, found that students incidentally learnt about 15% of the unknown words they met while reading. Students with higher school grades and reading ability learned more incidentally than those with less ability or in lower grades. There is no discussion on whether there is a point at which the ability is so low that no incidental vocabulary learning takes place. While the analysis showed that incidental vocabulary learning was small, the authors suggest that repetition through repeated reading (ER) can strengthen word knowledge. Table 2.1 lists second language vocabulary acquisition studies from reading. Nadsat words are ‘nonsense words’.

Table 2.1 Summary of studies of vocabulary acquisition from reading

Study	Total text / target words	Level	Reading time / test type	Gains
Pitts, White and Krashen (1989) Experiment 1	6,700 words 28 nadsat words	35 ESL learners	60 minutes MC test	1.81 words - 6.4%
Pitts, White and Krashen (1989) Experiment 2	6,700 words 30 nadsat words	16 ESL learners	watch video 40 minutes reading MC test	2.42 words – 8.1%
Zahar et al (2001)	2,400 / 30 words	144 grade 7 5 groups	1 hour (listen/read) Nation style (appendix xx)	<u>immediate post-test</u> 1.48 – 3.03 words / 7.5% <u>1 week post-test</u> little change
Tekmen and Daloglu (2006)	2,400 / 30 words	99 University 3 groups intermediate - advanced	1 hour (listen/read) Nation style (appendix xx)	<u>immediate post-test</u> 3.12 - 5.32 words / 14.0% <u>1 week post-test</u> little change
Day et al (1991)	1032 words / 17 words	92 high school and 200 university EFL	30 minutes MC test	<u>immediate post-test</u> 1 word – 5.8%
Horst et al (1998)	21,232 words 23 words	34 ESL	6 hours MC test and word association	<u>immediate post-test</u> MC, 5 words – 22% association 3.7 words – 16%

Waring and Takaki (2003)	5,840 words / 25 nadsat words	15 university	about 1 hour 3 tests, word recognition, MC, translation	<u>immediate post-test</u> recognition 15.3 – 61% MC 10.6 – 46% translation 4.6 – 18.4% <u>1 week delay</u> recognition 11.1 – 44.4% MC 7.9 – 31.6% translation 1.9 – 7.6% <u>3 month delay</u> recognition 8.4% - 33.6% MC 6.1 – 24.4% translation 0.9 – 3.6%
Brown et al (2008)	5,500 words / 28	35 university	1 hour meaning translation test MC test	<u>immediate post-test</u> translation 4.10 words – 14.6% MC 12.54 words – 44.8% <u>1 week delay</u> translation 2.34 words – 8.4% MC 12.46 words – 44.5% <u>3 month delay</u> translation 0.97 words – 3.5% MC 11.37 words – 40.6%

The fourth assumption to be accepted in support of vocabulary acquisition from ER is that acquiring a word is a cumulative process. An initial exposure to an unknown word may create a vague or partial understanding (Laufer, 2003) and the word needs to be repeated before it is forgotten, which is one of the fundamental premises of ER. Laufer (2005:3) states that 'picking up words from context has limitations unless learners are flooded with input'. The word may need to be repeated from eight (Horst et al, 1998) to twelve (Saragi et al, 1978; Waring and Takaki, 2003) times to enrich and strengthen the learners' knowledge of it.

Nation and Wang (1999) found that for a group of L2 adult readers with knowledge of 1487 word families, their reading series level introduced 459 new words. Of these new words, only 108 occurred more than 10 times in the 200,000 total, which would equate to 540 new words acquired from a million read. At lower levels, new words are more frequently recycled, but as one moves to texts with a larger number of words, new words become less frequent and more needs to be read to encounter them the required number of times. The difficulty of accumulating vocabulary through exposure and repetition is the large quantity of text that would need to be read for gains to be made. Based on these figures, the sample population for the current study (see 3.1.1) would need to read 83 readers (12,000 words each) to gain 500 – 600 words, which would take

them almost 3 years (10 days to read a book) of continuous reading. This would be a difficult task for students who have previously only read 2 – 3 books a year.

Cobb (2007) argues that ER cannot provide L2 learners with sufficient opportunities for acquiring the vocabulary necessary for the adequate reading comprehension of English texts. Cobb suggests that a minimum of six exposures to a word family (argue, argument) would be required for acquisition and that at least 3,000 word families would need to be acquired for adequate comprehension. Cobb (2007:41) estimated that in ‘a year or two a student could read roughly 175,000 words (about 14 readers of 12,000 words). By randomly selecting ten words at each word level and determining their frequency from 175,000 words in academic reading, fiction and the Press, Cobb found that the selected words did not occur six times or more at the 2,000 and 3,000 word level, thus drawing the conclusion that students would not be exposed to the words enough to acquire them.

McQuillan and Krashen (2008) summarized reading speeds of students of a variety of levels from eleven studies and concluded that with an average reading speed of 100wpm, this quantity of reading amounted to just 29.2 hours over the two years of language study, or 2.4 minutes of free reading per day (pg. 106). They suggest, therefore, that

three times as much reading could be given, which would ensure the necessary exposure to new words. Other studies of reading speeds, however, suggest 59wpm (Sheu, 2003 – junior high Taiwan), 68wpm (Bell, 2001, adults in Yemen) and less than 100wpm for advanced level learners (Jensen, 1986). Standardized at 60wpm over 200 school days would amount to about 14 to 15 minutes of reading per school day.

Three times the amount of reading, as suggested by McQuillan and Krashen (2008), would be 525,000 words, and at 60wpm this would equate to about 44 minutes of reading per day over 200 school days. The task could be manageable for beginner readers if time was made available each and every school day, but it may be less manageable for weaker H.K. students in higher forms who are examination-hardened veterans. These students may require considerable motivation to read a little every single day, particularly on top of several hours of normal homework per day.

2.2.3 Opposition to the Noticing Hypothesis

In opposition to the noticing theory and in agreement with Krashen, Cleeremans et al. (1998), from their studies in pure cognitive science, suggest that while awareness often appears to accompany learning and might often enhance it, it may not always be necessary for learning to occur. Truscott (1998) agrees, arguing that noticing is helpful

but not necessary and that research on form-focused instruction and feedback suggests that awareness of form is not only unnecessary but also unhelpful. He states that noticing is necessary for the acquisition of metalinguistic knowledge but not competence. Truscott (1998:107) argues that attention research, cited in support of the Noticing Hypothesis, indicates that language acquisition requires little more than global awareness of input, which virtually no one would dispute (pg. 104). Truscott (1998:117) does note, however, that his arguments apply to grammar and not to formal instruction in general and that the Noticing Hypothesis may be applicable in other areas, such as lexical learning.

A study of 44 University ESL students compared their ability to notice unknown words and formulaic sequences (Bishop, 2004). Unknown words were clicked on (computer mouse) significantly more often than unknown formulaic sequences. However, making formulaic sequences typographically salient (using colour and underlining) significantly increased the number of times the formulaic sequences were noticed (clicked on). Typographical elaboration, which aided noticing, also appeared to increase global comprehension of the text. While the study made no distinction between comprehending a lexical item in context and actually acquiring the item, it does suggest that some lexical items, in this case formulaic sequences, may need to have learners'

attention drawn to them to highlight that the item is actually ‘problematic’, to use Schmidt’s term (Schmidt, 1990:129).

Kim (2006b), in a study of 297 Korean university students, found that typographical enhancement on its own did not aid form and meaning recognition of vocabulary, but both explicit and implicit lexical elaboration did aid meaning recognition of vocabulary. Reading a text with explicit lexical elaborations seemed to be more conducive to L2 learners’ recognition of the meanings of unknown target words than reading a text with implicit lexical elaborations (explicit elaborations: naming, questioning and, in this study, synonyms and definitions $x = y$, implicit elaborations – apposition x, y). The study suggests that simply drawing attention to an unknown word alone will not aid acquisition of the word. As noted earlier, a text may provide insufficient cues to enable guessing or the reader may guess incorrectly. This study suggests that some type of reading with tasks (R+) where explicit elaboration can be directed toward targeted lexical items might be more effective in the H.K. context.

Lee (2007b), however, found that while using textual enhancement to aid the learning of target forms (passives) among 259 Korean 17-year old students had a positive effect, there was an unfavourable effect on comprehension. Lee found that when topic

familiarity was highlighted, students' comprehension of the article was enhanced but they did not attend to the target grammar (passives), but if the grammar form was highlighted, it was enhanced at the expense of comprehension. Lee's conclusion suggests that attention to formal aspects of the text undermined attention to other aspects (meaning). Topic familiarity aided comprehension to such an extent that meaning was clear without the need to focus on the grammar. If students are entirely focused on meaning, as with ER, and are reading self-selected books on familiar topics, they may have no need to attend to specific grammar items or to unknown vocabulary that does not hinder meaning.

Kim (2006a) notes that reading for comprehension (semantic processing) and reading for acquisition (syntactic processing) are different. Meaning is processed by L2 readers before they process for form, so only if a text is easy or sufficiently enhanced and meaning is gained easily will focus on form take place. Kim (2006b) argues that because reading is meaning-focused, it cannot produce native-like competence in learners' output, such as in writing. VanPatten (2002) agrees that learners process lexical items before grammatical items and the grammatical structures required for meaning (-ed) are processed before the more redundant items (subject verb agreement). This would suggest that grammar and writing skills would not be easily acquired

through ER, although a number of ER studies have shown that this is possible (Mangubhai, 2001; Tsang, 1996), perhaps where the semantic processing was less challenging, enabling free mental resources for processing syntax.

The argument on textual elaboration and implicit or explicit elaboration brings us back to the discussion of implicit versus explicit learning. There are relatively few studies that directly compare implicit and explicit learning. Norris and Ortega (2000:521) list just 14 out of 77 reviewed, but none specifically involve reading.

The studies generally showed an advantage for the explicit learning groups. It should be noted, however, that nearly all the studies were quite short, some less than 2 or 3 hours, and the longest (DeKeyser, 1997) only 12 weeks. The studies would seem to favour explicit learning, as implicit learning takes longer to develop.

Doughty (2003) notes that after reviewing cases for and against L2 instruction, the weak evidence is that instruction is potentially effective providing it is relevant to students' needs. Relevance is the challenge for any language instructor and it could be concluded that relevance is just as vital for implicit learning and possibly more so as the students need to have their attention held for a long enough period of time for implicit

acquisition to have an effect. ER has the ability to meet students' needs, at least on an interest basis, as they self-select the materials. In Hong Kong, however, it may not be seen to meet students' immediate goals, which are related to passing examinations (Morris, 1985; Dykes, 2007).

Nation (2001) concludes that learners should not rely solely on incidental learning from context, but that direct vocabulary learning and incidental learning are complementary activities. Independent reading with vocabulary tasks, however, can be particularly effective by enabling learners to encounter words in a meaningful context (Rosszell, 2007).

One of the more influential theories to argue for the implicit acquisition of language is Krashen's theory of second language acquisition. Krashen's (1976, 1981, 1982, 1985) theory is one of the most comprehensive in its claims and it provides some of the 'greatest' theoretical support for ER as a means of promoting second language development. Essential to his hypothesis is the theory that language can be acquired subconsciously.

2.2.4 Implicit learning: Krashen's Comprehension Hypothesis

The Comprehension Hypothesis, as it is now called (Krashen 2004a), consists of five sub-hypotheses: 1) the Acquisition-learning Hypothesis, 2) the Natural Order Hypothesis, 3) the Monitor Hypothesis, 4) the Input Hypothesis and 5) the Affective Filter Hypothesis. This paper deals primarily with the three hypotheses related to uninstructed reading and language acquisition (1, 4 and 5), although it is acknowledged that the five hypotheses are interrelated.

Krashen summarized his sub-hypothesis by suggesting that if a child is allowed the necessary input, complete competence in the target language (first or second) appears to be inevitable (Krashen 1976:163). Krashen and Terrell (1983) claim that acquisition depends crucially on the input being comprehensible and comprehensibility is dependent directly on the ability to recognize the meaning of key elements in the utterance. Thus, acquisition will not take place without comprehension of vocabulary (pg. 155).

Krashen distinguishes between learning and acquiring a language and sees the two as distinctly different ways of developing ability in a second language. Acquisition takes place subconsciously in an identical process to that utilized by children in acquiring

their first language. Krashen believes that a second language can be acquired in the same manner as the first on a sub-conscious level (Krashen, 1989). Krashen makes a clear distinction between the conscious and subconscious processes. He argues that when a learner focuses on the 'form' of the language, then the process is one of conscious learning, but when the focus is on the 'message' of the language then the process is one of subconscious acquisition. 'Learning', on the other hand, refers to 'conscious knowledge of a second language, knowing the rules, being aware of them, and being able to talk about them' (Krashen, 1982:10).

Krashen (1989:454) states that 'conscious language learning does not appear to be as efficient as acquisition from input'. His recommendation, therefore, is that L2 students should have as much access to comprehensible input as possible for implicit learning and do not require explicit rules and systematic practice of these rules.

The Acquisition-Learning Hypothesis is one of the key theories in explaining the role of ER in language acquisition. Eskey (2005) notes that the best way to acquire the extensive vocabulary required for reading widely in a second language is through reading and that the process of reading is largely invisible and occurs at the

subconscious level. ER, which utilizes a learner's desire to understand a message, is an acquisition-rich environment.

Gregg (1984) had a number of criticisms of Krashen's hypothesis and the two main ones are summarised here. The first objection is for learners who learn language only in a formal 'instructed' setting where instruction is in the first language rather than the target language. According to Krashen's theory, these learners would never speak the L2. Gregg states that he learnt Japanese in a very controlled, formal manner but speaks the language well (pg. 81).

Krashen (1989:454) has softened a little in his non-interface position in that he acknowledges that learning has some impact on language development. He admits that 'while a substantial part of these gains is probably due to incidental comprehensible input, there are gains over and above what one would expect from comprehensible input alone'.

The second objection to Krashen's theory is to do with falsifiability. Krashen does not provide any evidence that learning and acquisition are indeed two separate systems; neither does he provide a means for determining this, so the claim is unable to be

evaluated. Both Gregg (1984:94) and McLaughlin (1987:56) suggest that Krashen's theory has serious flaws in that terms are ill-defined or undefinable and the empirical basis of the theory is weak.

Ellis (2001) argues for an interface position between acquisition and learning and suggests that explicit learning may indirectly affect the acquisition of implicit knowledge by focusing learners' attention on features of input which are critical for language to be acquired (see 2.2.1). A number of studies have concluded that both instruction and exposure to language contribute to the enhancement of language proficiency (Long, 1983).

The Comprehension Hypothesis is central to Krashen's overall theory of acquisition and a key theory in support of ER. This theory attempts to answer the key questions in language acquisition and language education: How does language acquisition occur? Krashen (2003:4) suggests that 'we acquire language in only one way: when we understand messages; that is when we obtain "comprehensible input". Krashen (2004a) actually changed the name of the hypothesis from the Input Hypothesis to the Comprehension Hypothesis to emphasize that mere input is not sufficient.

The Comprehension Hypothesis is a supplement to the Natural Order Hypothesis, which states that a learner will acquire language in a natural order. The theory states that if our current linguistic level is at 'i', then the next structure we acquire is $i + 1$, which is the next level along the natural order. We are able to understand $i + 1$ with the help of previously acquired linguistic knowledge as well as our knowledge of the world and the situation, all of which we call 'context'.

Krashen (1989) assumed a language acquisition device (LAD): that is an innate mental structure capable of handling both first and second language acquisition. Input such as $i + 1$ is capable of activating this device. The implication for UR or ER is that if sufficient comprehensible input is provided then language is automatically acquired. The language teacher has no need to deliberately teach the next structure in the natural order, as it will be automatically provided and thus acquired at the appropriate time.

The main criticisms of the hypothesis are that Krashen does not define what 'sufficient' input is. He is also not specific as to what level of knowledge is (i), thus making $i + 1$ indefinable and making the theory unfalsifiable (Gregg, 1984; Ioup, 1984).

On a daily basis, however, classroom teachers define 'i' by ensuring that they give students work at the 'appropriate level' in whatever subject they are teaching. In reading,

an educator attempts to define an individual student's reading level through the use of various measures such as running records (Holdaway, 1979) or vocabulary tests (Nation, 2001). Indeed, Carver (1994) argues that for learners to increase their vocabulary size, they need to read material that is not too easy for them, as while easy reading may increase depth of vocabulary knowledge, it is unlikely to increase the breadth of vocabulary knowledge. In order to do this, one must be able to ascertain the learner's current 'level'. A number of researchers (Fry, 1981; Nation, 2001; Koda, 2007) recommend that for meaning-focused input, learners need to have 95%-98% coverage for learning vocabulary and 98 - 100% for fluency development. On a practical level, it is suggested that this is $i + 1$, which is definable for vocabulary. Generally, when selecting a reader for a student at an appropriate level, it is the vocabulary that we are concerned with rather than the grammar of the book.

If an educator cannot determine a learner's reading level (i) then often the learner themselves can. With a little training, they can fairly quickly determine whether a book is too difficult to read and locate one at their appropriate level ($i + 1$). Vocabulary levels of $i + 2$ or $i + 3$ could be defined as lower and lower percentages of known words equating to lower levels of comprehension, with $i + 0$ meaning that all words are known and improved comprehension (Hu and Nation, 2000).

Krashen argues for an Affective Filter, which is essentially a block which shields the LAD from the input necessary for acquisition. The filter is what allows comprehensible input to be comprehended. If a learner is anxious or has low self-esteem, he or she may understand the input but it will not reach the LAD. A block, the Affective Filter, will keep it out. Krashen (1982:31) says that ‘those whose attitudes are not optimal for second language acquisition will not only tend to seek less input, but they will also have a higher or stronger affective filter’. It is not clear how low the affective filter must be before it can facilitate language acquisition.

Gregg (1984), in his criticism of the Affective Filter, gave the example of a Chinese native speaker who, while having near native-like English, had not acquired the third person singular – s. In Krashen’s view this would be due to the Affective Filter, but as Gregg notes, there is no explanation as to how the Filter could let in most input and yet filter out the third person singular (pg. 85).

In extensive reading, reading is, by definition, individualized, with students choosing the books they want to read. There are no tests or assignments and the objective is ‘obtaining pleasure from the text’ (Susser and Robb, 1990). If there is an Affective Filter, as described by Krashen, then it should be at its lowest with an extensive reading

programme providing little pressure (in theory). In the H.K. context, however, with lower ability L2 students who do not have a reading habit, the sight of an actual 'book' as opposed to short passages may in fact be quite daunting, particularly if the students, upon completion of the book, know that they will be required to complete a report. Weaker readers who have had many years of L2 classes but still have limited vocabulary and comprehension may lack self-esteem or have a poor attitude to learning English, thereby having a higher affective filter, making acquisition from ER difficult. In this study students are given a book to read with simple reading comprehension questions. For students, who do not have a reading habit, the affective filter may have been quite high which may, in part, explain why some student did not read the first book in this thesis (Table 4.2).

Elley (2001:129) notes that when the conditions for the comprehension hypothesis are met and comprehensible input is abundantly available, the input is slightly above students' current level of competence, and as the student 'repeatedly focuses on the meaning of a large number of interesting messages...he or she incidentally and gradually acquires the forms in which they are couched'. The problem in H.K. is ensuring that text is interacted with.

Lai (1991) carried out a study involving 11 - 13 year olds from eight H.K. secondary schools in an ER study (n = 1133) specifically to investigate the validity of Krashen's input hypothesis, as it was then called. The students were matched to control and experimental groups. One school was ranked as having a high level of English, two as mid-level and two as low-level. The control group had two lessons per week for reading, using four to six set books. The experimental group had one lesson for the same task and the other lesson was used for ER tasks. Lai notes that in this lesson (40 minutes) the students returned the borrowed books and completed comprehension question cards, marked the answers and filled in a results wall chart. They then borrowed another book to read at home. It would seem that little reading took place in the lesson itself. The programme was carried out for one year. Out of the five schools involved in the first study, only one experimental school showed an advantage over the control: a mid-level Chinese medium school. In language proficiency, the two low-level schools both performed slightly worse than their control schools, with small non-significant increases for the other mid-level school and the high level school.

Waring (2001) suggests that there is an advantage in this type of study for lower ability learners, as they are capable of greater gains. However, the lowest ability students failed to make gains, which suggests that perhaps there is an optimal language level required

for language acquisition to occur through ER, as already noted by a number of writers (Nuttall, 1996; Coady, 1997; Liu and Nation, 1985; Laufer, 1997). Students whose language is too weak may struggle to comprehend text sufficiently through unguided reading to make sufficient gains, while those at a higher level may need to read a great deal in order to meet new vocabulary at sufficient frequencies to be retained.

In reading comprehension, the same mid-level school made a significant improvement, with one low-level school and the high-level school doing worse. In vocabulary recognition, the same mid-level school and both low-level schools report a significant improvement over the control groups. The high-level group again did worse than the control group. While the study shows that those who read more showed a corresponding improvement in language ability, the trend did not apply to the students who read the most, particularly in the high-level school. Teachers reported that these students were in the habit of skimming through books quickly in order to get enough information to answer the multiple-choice questions. It is surmised that completion of a book would result in an additional check (credit) on the wall chart. Some schools award prizes and certificates for the completion of a certain number of books, which may account for the higher ability students reading a large number of books without making any real language gains. Lai concludes that the study only provided partial support for Krashen's

Input Hypothesis, since it was found that exposure to comprehensible input did not necessarily guarantee acquisition. Lai concluded that depth of processing was an important factor in determining whether acquisition took place. There may be a role for ER with average ability students from poor language environments; however, if a school-wide or nationwide reading programme is to be implemented, it needs to be able to meet the needs of students at all ability levels, including the weakest. The study does not answer the question of whether the lack of a noticeable language gain was due to the lack of engagement with the task (reading) or an inability to acquire language from reading. There is a suggestion that some students skimmed the books and were therefore not sufficiently engaged. This would suggest that the input hypothesis was not effectively tested. In this thesis, subjects indicated their level of engagement in a single book providing a better measure of the quantity of input in relation to language (vocabulary) gains.

Krashen (1989:652) reviews 144 studies in his attempt to provide evidence for the Input Hypothesis through ER. He notes, however, that nearly all of the research reviewed is from first language acquisition, but suggests that the small amount of second language research points in the same direction. The small amount of research he quotes at this

time consists of only three studies, which may suggest a role for ER in second language acquisition but hardly confirms such a role.

Fifteen years later, Krashen (2004b) continues to argue for language acquisition through ER and summarizes 54 studies of extensive reading (1). In 51 out of 54 comparisons (94 percent), readers did as well as or better than students who engaged in traditional programmes. A finding of ‘no difference’ between free readers and students in traditional programmes suggests that free reading (ER) is just as good as traditional programmes.

Table 2.2 Study duration and effect of ER on language acquisition, Krashen (2004b:2)

Duration	Positive	No difference	Negative
Less than 7 months	8	14	3
7 months – 1 year	9	10	0
Greater than 1 year	8	2	0

Of these 54 studies, however, only nine involved second language learners showing a positive difference for English for ER (Aranha, 1985; Mason and Krashen, 1997 study 1; Elley, 1991 – p1 survey; Elley, 1991 – Singapore n = 512; Elley, 1991 – p3 survey;

Mason and Krashen, 1997 study 2 – four-year college student study; Mason and Krashen, 1997 study 3 – two-year college student study; Lituanas et al, 2001; Elley and Mangubhai, 1983). The three Elley studies were all part of the Singapore Reading and Acquisition Programme (REAP). Only five studies were for ER or uninstructed reading: Elley (1983), Mason and Krashen (1997 – 3 studies) and possibly Aranha (1985 – this study is not reviewed, as few details are given as regards the structure of the reading programme to determine whether it was guided reading or pure ER). The remaining four studies (Lituanas et al, 2001 and Elley REAP studies) were shared-book reading which means reading with considerable time spent on supplementary activities and/or teacher guidance or some form of tasks after reading (R+). Krashen in fact only offers four pure UR studies in support of his theory, three of which were by himself and involved university students (Mason and Krashen, 1997 – 3 studies; Elley and Mangubhai, 1983). Krashen's analysis shows a paucity of UR studies for L2 learners prior to 2004 to support his comprehensible input hypothesis, particularly for school-aged learners.

Iwahori (2008) comments on a lack of studies for secondary school students demonstrating benefits for ER. He lists only two for reading comprehension / vocabulary (Lai, 1993b; Sheu, 2003), both of which are discussed in this review.

Day (2009) lists 21 studies in support of extensive reading, of which 13 relate to vocabulary or general language proficiency. Nine of the studies are discussed in some detail in this thesis: Elley and Mangubhai, 1981 (2.3.5); Pitts et al, 1989 (2.2.2); Hafiz and Tudor, 1990 (2.2.5); Elley (1991), in fact, refers to a number of book flood studies (R+), which employed shared book reading and are discussed in 2.3.5 and Appendix B2; Lai 1993a (2.1.3) and 1993b (1.5); Bell, 2001 (2.3.4); Sheu, 2003 and Iwahori, 2008 (2.2.5).

The four studies from Day (2009) not reviewed in this paper are not discussed primarily due to space and lack of ability to generalize from the results to the general population. In Robb and Susser (1989) and Masuhara et al (1996), the control and experimental groups had unequal exposure to input, so the experimental group with the greatest exposure outperformed the control group. Cho and Krashen (1994) employed a sample of only four students reading a single book, which showed an acquisition rate of about one word for every 8,000 read, and Horst (2005) was largely a feasibility study on measurement in ER.

2.2.5 Extensive reading studies

It can be extremely difficult to carry out extensive reading studies over an extended period of time due to the difficulty of controlling both internal and external variables. In spite of the commendations for ER in the literature, there are relatively few good pure ER studies and very few involving comparisons of ER with alternative language-rich programmes. No quantitative studies have attempted (to date) to compare ability groupings within a single class or sought to deliberately determine whether ER is suitable for L2 learners of all ability levels. The studies selected for review are those that have been commonly cited by others, studies involving students from an L1 which does not use a Romanised script and studies which enable some generalisations to be made for the sample population of this study. The review has only included studies that involve ER in English.

In possibly one of the most quoted ER studies, Hafiz and Tudor (1989) report on a study involving ESL students of Pakistani backgrounds aged 10 – 11 living in the UK. While the study is oft cited as an L2 study, the subjects were possibly bilingual, as most had been born and educated in the UK. A programme which involved the students (n = 16) taking part in approximately sixty additional hours of reading-class time over a period of twelve weeks showed significant gains for the experimental group over two control

groups (n = 15) in a general reading proficiency test of comprehension and vocabulary items. There were no controls over the exposure the subjects had in their regular classes or from the environment, which makes the actual gains from ER difficult to quantify.

In another similarly well-quoted study by Hafiz and Tudor (1990), a group of 25 male 15-16 year old Pakistani students were matched with a control (n = 24). The experimental group had four hours of classes plus four hours of silent reading per week (total 90 hours) while the control group had no English input at all over a twenty-three week period. Initially the study employed two control groups, but data from one of the groups had to be discarded because of 'a lack of commitment to the project on behalf of the learners involved' (Hafiz and Tudor, 1990:33). Hafiz taught the experimental group, although he had not been a member of the school's staff prior to the experiment. Hafiz and Tudor admit that 'the novel effect of the programme itself, combined with the personal contribution of the experimenter, may have constituted significant causal factors in the overall success of the project' (pg. 41). It is suggested, however, that a novel effect may not be 'novel' enough to last for twenty-three weeks, although the early gains could have boosted the averages.

While the experimental group did make considerable language gains in writing and vocabulary, these gains may have been due to the language classes, ER or a

combination. The study simply shows that students will gain language if they are exposed to it.

In Mason and Krashen (1997 – Experiment 2) the study involved two classes from the English literature department at a university in Japan and two junior college classes (n = 128). All were tested on a 100-item cloze test as per experiment 1 and divided into experimental and control groups (see 2.1.3). The same criticisms for the cloze test as noted for experiment 1 also apply with this experiment. The control classes had regular lessons focusing on direct teaching of reading comprehension and intensive reading of short, difficult, assigned passages, which required considerable dictionary work. The experimental groups read books, completed book summaries in English and wrote appreciation notes about the books in Japanese. The study ran for a year and both experimental groups made significant gains in the post-test at the conclusion of the study. The study does not report how much time was actually devoted to reading or the quantity of books completed.

Experiments in ER can be particularly difficult to carry out in real classroom settings, as there are numerous variables that need to be controlled. The researcher needs to make every effort to control them if possible or to report them faithfully if they were not

controlled. In this study, the experimenter only taught the experimental classes, which could affect the results, but this receives no comment. The results are confusing, as the reported degrees of freedom (*df*) do not match the sample size (*df* for independent samples t-test is $n-2$ – in the university group, $n = 79$ and the $df = 72$, with similar errors for the school sample). If one cannot have total confidence in the statistics, then the results are less reliable. These studies both compare input-rich extensive reading with less-rich traditional methods and show an advantage for ER. The studies are unable to comment on how ER compares with alternative language-rich programmes.

Sheu (2003) studied 98 Taiwanese grade 2 students aged 13-14 years old who had had less than one year of previous English study. One experimental group read graded readers for 45 minutes a week in class and the other read books suitable for native English-speaking children (BNESC). The control group spent the time revising grammar work from previous lessons. The study does not mention a time frame. The pre-post test was the Cambridge Key English Test (KET) with a total of twenty questions (six matching vocabulary question, seven MC grammar questions and seven true-or-false comprehension questions). It is somewhat surprising that this study is reported as significant (Day, 2009; Iwahori, 2008) with such minimal testing. Both experimental groups showed a significant improvement in reading comprehension on a

one-way ANOVA. Attitude scores for the BNESC group showed a significant decline, possibly due to the difficulty of these types of books for beginner L2 students.

A similar study to the one above was carried out with second-year Taiwanese secondary students (Sheu, 2004). The study used a 20-question pre-post test. Modifications were made from the first experiment, which included a strategy training session at the beginning of the experiment for the experimental group and the introduction of a wall progress chart, group discussions for sharing reading and teachers' reports for 'demonstrating their participation and for encouraging students to read' (pg. 221). The study involved the experimental reading group in two 45-minute lessons a week for four months. The two experimental groups achieved a statistically significant level of improvement in all three language areas. The experimental groups showed a significantly improved attitude towards reading and learning English.

The results of the two studies together suggest that just giving students books to read may improve language acquisition but may not positively impact on learners' attitudes (study 1). However, the introduction of collaborative activities, while aiding language acquisition, may also improve learning attitudes. Aiding reading assisted the BNESC

group particularly by making their reading more comprehensible and the tasks may have lowered the affective filter (Krashen, 1989) for both experimental groups.

Iwahori (2008) carried out an extensive reading study in a Japanese secondary school (age = 16 – 17 years) for seven weeks ($n = 33$). The students are described as beginners who had studied English for 4 – 8 years previously ($M = 7$). Students were provided with a large number of graded readers and comics, which they were encouraged to read at home over the four school weeks and three holiday weeks of the study. While a target of 28 readers was set, there is no record of the mean number read. Students had agreed to join the class and signed a consent form that explained the study. They were therefore aware that they were part of a study and this, together with their choosing to take part, may have resulted in quite high motivation, which could be sustained due to the short length of the study. Students were measured on a cloze test, which measured general language proficiency. There was no control group. Results show a significant improvement on the cloze post-test ($p = .003$), although the actual mean gain is quite small at 3.42 over a 100-item test ($r = .49$, not provided, see 4.0). The same cloze test was used pre and post and students may have had some familiarity with it. The results, while suggesting improved general proficiency, give no clue as to improvements in spelling, grammar, vocabulary or text structure. Iwahori describes the results as ‘less

robust' (pg. 83). Without a control group, it is not possible to claim that an improvement in the dependent variable is solely due to the treatment: reading.

Nishizawa et al (2010) carried out a study on 37 weak university engineering students in a Japanese university. The students, aged 20 – 22 years old, read independently for 45 minutes a week for 120 weeks over four consecutive academic years. The study showed a strong correlation between their TOEIC scores and the amount of reading. The TOEIC scores recorded significant gains compared with comparison groups. The most critical factor for success was the quantity of reading completed. There was little improvement for reading less than 100,000 words and a threshold level of 300,000 for improvement was noted. The study showed that students who are weak and reluctant to study in English could improve their language ability without sacrificing their primary (engineering) study. A second critical factor for success was starting with very simple stories such that students could read without the need for concurrent translation. In spite of the success of the programme, demonstrated through statistical averages, 25% of the students had read less than the threshold level of 300,000 words after three years and 16% after four years. There is no information as to why this group of students failed to engage in reading. The mean words read for the bottom nine students after four years was 310,000 words (only about eight 10,000-word novels a year) but the students still

scored 15 points higher than the national average on a TOEIC test. The improvement seems to have been possible without reading very extensively. The challenge still remains, however, of encouraging H.K. non-readers of L2 to read without assistance (UG) eight books a year and to maintain that momentum for four years or to read thirty books in a single year when in the previous year they read two or three. It is unknown how the results for this group of university students may relate to H.K. secondary students.

2.2.6 Summary of Part 2

This section has compared implicit with explicit learning and suggests that a combination of implicit and explicit learning, particularly of vocabulary, may be more effective than either method in isolation. Implicit acquisition of vocabulary from UR requires a reader to notice an unknown word, guess its meaning and retain the word for some time, during which time it is encountered again and has its knowledge strengthened. A number of problems were outlined for these assumptions. Krashen's comprehension hypothesis was discussed as a key theory in support of language acquisition from UR, although he offers little empirical support for L2 learners. The section concluded by examining studies of UR not already covered in the thesis. There

are few studies in support of ER and in particular studies which compare UR in the form of extensive reading with other language rich programmes.

The inability of some readers to infer correctly and thus to acquire vocabulary from unguided reading has led some second language acquisition (SLA) researchers to describe the learning of vocabulary from reading as haphazard and inefficient (Haynes, 1993; Laufer and Sim, 1985; Paribakht and Wesche, 1997). H.K. extensive reading studies have suggested the relative ineffectiveness of UR as a means to improve language, particularly vocabulary and reading comprehension. Hulstijn (1992), from several adult L2 studies, concluded that in true incidental learning tasks, retention of word meanings was very low. Wesche and Paribakht (2000) found that while for adults in a university setting studying English, reading for many did result in significant acquisition of L2 vocabulary, direct instruction led to the acquisition of an even greater number of words as well as greater depth of knowledge.

While there are some strong claims in favour of language acquisition from reading, the claims are less well supported in L2 than in L1, particularly in the H.K. context. The claim that vocabulary acquisition from uninstructed reading is haphazard and inefficient and that this type of reading, particular where it is practised in an extensive reading

programme, does not pedagogically fit the H.K. system and has largely failed, led the researcher in a search for alternative reading methods to aid vocabulary acquisition. In countries where reading was introduced to aid or boost local programmes, it was often introduced in the form of shared reading or reading with tasks (R+) rather than as pure ER.

2.3.0 Reading with tasks (Part 3)

Part three examines the effectiveness of reading with the use of supplementary tasks (R+) as an alternative to UR. The section discusses Swain's emphasis on the role of output and then discusses R+ in its role in following an interactive reading model. R+ is seen as an effective means of developing students' schema and the importance of this is discussed. A key component in reading with tasks is the type of tasks employed. Section 2.3.4 examines the involvement load hypothesis and supporting studies.

Researchers such as Williams (1986:42) and Day and Bamford (2002:138) are quick to suggest that 'reading is its own reward' and the learners' experience of reading the text is at the centre of the extensive reading experience. For this reason, they suggest that extensive reading is not usually followed by comprehension questions because it is a complete experience in itself. Day and Bamford (1998:140) argue that if readers are

required to do exercises to demonstrate comprehension, or to practise vocabulary or reading skills, it confuses or distracts from the reading purpose. However, their use of the term ‘extensive reading’ is rather confused, as even while they were suggesting that ER should take place without follow-up activities, they edited a book in 2004 entitled *Extensive Reading Activities for Teaching Language*.

Krashen (2004b) suggests that the forms of supplementation that are most useful are those that help make input more comprehensible, more interesting, and encourage students to obtain more input. Supplementation that research has shown to be not effective are those that attempt to use means other than comprehensible input. Tasks such as reaction reports may be ideal for checking that reading has taken place but may not actually provide a learner with any input, or be interesting to write. While most supplementary reading tasks involve some form of writing activity, from the answering of questions to the writing of summaries, oral activities are sometimes used, particularly in shared book reading (Elley and Mangubhai, 1983: see section 2.3.5 and Appendix B2). Oral tasks in a group setting can provide others with input and be interesting, or even entertaining, and may encourage other group members to seek input (read the book).

In contrast to the pure ideals of ER and to Krashen (1989, 2003), who suggested that comprehensible input is the essential element necessary and sufficient for language acquisition, Gass and Selinker (2001:277) state that 'input alone is not sufficient for acquisition'. They argue that with input, syntax is not necessary for comprehension. For example, with some basic world knowledge, the words 'bit', 'girl' and 'dog' could only be interpreted one way. Gass and Selinker suggest that with language production or output, however, one is forced to put words into some order. It is suggested that a reader will never see the words 'bit', 'girl' and 'dog' out of order without the appropriate prepositions and articles and that constant exposure to the correct form through reading, as suggested by Elley (2001), will lead to language acquisition and the ability to use it.

Swain (1985: 248) suggests the idea of learners 'negotiating meaning', which is not just getting a message across but being pushed to convey a message coherently, precisely and appropriately, may force learners to move from semantic processing to syntactic processing. Swain (1985) believed that what was lacking in students' development was the opportunity to use language productively. Output had traditionally been seen solely as a means to practise what had already been learnt. Swain's (1985) idea is that if learners are 'pushed' or 'stretched' in their desire to be understood, they might modify what they say in order to improve comprehension or they might try out forms they had

not used before. Swain and Lapkin's (1995:371) claim for output is not to minimize the role of comprehension or input but to make the case that 'sometimes, under some conditions, output facilitates second language learning in ways that are different from, or enhance, those of input.'

The noticing or triggering function (Swain, 1985, 2005) suggests that as learners attempt to produce language, they may notice that they do not know precisely how to produce the meaning they want. This may bring the learners' attention to something they need to discover, possibly directing their attention to relevant input. Output may affect the acquisition of implicit knowledge by focusing learners' attention on features of input that they see as required or necessary similar to the weak interface position of explicit learning affecting the acquisition of implicit knowledge.

A study by Izumi (2002) found that those students who engaged in output-input activities outperformed, in learning gains, those who were only exposed to input for the sole purpose of comprehension. The study showed that output did not just affect the subjects' output production strategies but also their L2 knowledge. Izumi (2002:569) suggests that output triggered deeper and more elaborate processing of the form,

whereas input enhancement may have caused more rehearsal at a relatively shallow processing level.

A study by Webb (2009) of Japanese second language learners found that productive learning of word pairs led to larger gains in both receptive and productive knowledge of orthography, and productive knowledge of syntax, meaning, and grammatical functions, while receptive learning led to larger gains in receptive knowledge of meaning. The findings suggest that if only one method is used, productive learning might be the more effective. Griffin and Harley (1996) note that productive learning typically results in more and stronger knowledge than receptive learning. Nation (2007) suggests that output sets up learning conditions that are qualitatively different from those of input, making both output and input important for learning, thus making a case for R+.

It is therefore perhaps not a case of learning language implicitly or explicitly but rather, as Doughty and Williams (1998) note, that knowledge (linguistic input) can be represented either implicitly or explicitly along a continuum and that both may contribute to language learning. Koda (2007) notes that while quantity of input is important, quality of input processed determines what emerges from learning (2.1.3), or as Lai (1991) suggested, ‘depth of processing’, which suggests that enhancing input, or

at least drawing attention to it, may be important for acquisition in spite of Krashen's (2004a) suggestion that ER alone is more efficient.

Krashen (1998) suggests three arguments against Swain's comprehensible output hypothesis. He suggests that comprehensible output is too scarce to make a real contribution to linguistic competence, that high levels of linguistic competence are possible without output and that there is some evidence which suggests that students do not enjoy being 'pushed' to speak. Krashen (1998) quotes a number of studies that show increased anxiety on the part of students who are pushed to speak.

With regard to speaking activities, Day and Bamford (1998:35) conclude that, as attractive and motivating as these activities may be, 'students should talk less about what they read and instead use the time for reading more'. They believe that reading is a much more effective means of acquiring language than speaking, and thus that any time spent on any form of output is time away from reading and acquisition.

It is suggested that output could play two other roles in the language classroom, however, by indirectly aiding acquisition. Firstly, when students talk about books they have read, share ideas about books or role-play characters, it may instil in others an

excitement to want to read the books and to share in the same experiences, which Krashen (2004b), noted earlier, suggested was a useful form of supplementation if it encourages students to obtain more input.

Secondly, therefore, the output of one student is input for another and the output is often adjusted so that it is comprehensible (Swain and Lapkin, 1995:373). Manning and Manning (1984) note that students recorded greater gains with ER when they discussed their reading with each other in pairs or small groups.

Morgan and Rinvulcri (1986) suggest that, 'sharing and discussing one's learning is far more motivating and far more memorable than grimly working on one's own' (pg. 6)....'it is the talking about words that anchors them and makes them permanent' (pg. 10). As a result of such interaction, learners gain insight from each other, and their ability to comprehend, retain, and use the words is enhanced (Morgan and Rinvulcri, 1986; Hall, 1992; Corson, 1997)

Atay and Kurt (2006) compared reading followed by two-hours a week of interactive tasks with reading followed by two-hours a week of discrete language tasks over a six-week period with a group of 62 Turkish eleven-year-old beginners who had had two

years of English instruction. The results showed an advantage for reading followed by collaborative interactive tasks, which substantially enhanced learners' motivation, aiding their vocabulary knowledge. Not only did the study show a gain for the 26 targeted words, but the experimental group also showed a gain on a general Cambridge standardized test that measured receptive and productive word knowledge.

One of the greatest difficulties in the Hong Kong classroom is being able to provide for meaningful and comprehended output in a class of 40+ students. Oral output, however, is in fact only one small aspect of an R+ programme, which is centered on enhancing input through the use of reading supplemented with a variety of tasks.

2.3.1 Aims of R+

Reading with tasks (R+), for the purposes of this thesis, is the reading of a complete graded reader (8,000 – 15,000 words) with a variety of written, oral, reading and / or listening tasks involving receptive and productive language in order to enhance language acquisition. Language acquisition is enhanced as the tasks can make the reading more comprehensible, recycle language and aid motivation. The tasks would generally be employed before reading, during reading and after reading. R+ is similar in design to the shared book method except that every student would have their own copy

of the reader and mostly read it on their own (Appendix B2). Reading with tasks (R+) may be capable of both initially attracting a learner through pre-reading tasks and the interest of the text and of directing attention through tasks to salient language points. R+ may allow many of the general benefits of reading (1.3.1) while providing students with structured assistance on various language aspects introduced in context through a book. Essentially R+, like the shared book method, is language presented implicitly and intentionally and may not necessarily involve direct instruction but rather provides opportunities to notice, use and explore language.

Table 2.2 (adapted from Waring, 2002) positions reading with tasks (R+) between totally free reading of an extensive reading programme (ER) and a totally teacher-controlled intensive reading programme (IR) where IR is the detailed analysis of short texts.

Table 2.3 A summary of the basic characteristics of ER, R+ and IR

	Extensive Reading (ER)	Reading with Tasks (R+)	Intensive Reading (IR)
LINGUISTIC FOCUS	Fluency, skill forming	Fluency, skill forming, attention to vocabulary, some analysis	Analysis of the language
DIFFICULTY	Very easy	Very easy – pre reading and tasks make it easier	Usually difficult
QUANTITY	A book a week	A book every 10 days	Little
SELECTION	Learner selects	Teacher selects	Teacher selects
WHAT MATERIAL	All learners read different things (something interesting to them)	Learners' ability grouped. A group read the same material.	All learners study the same material
WHERE	Mostly at home	Mostly at home, activities and discussion in class.	In class
COMPREHENSION	Checked by reports / summaries / conferencing	Checked by a variety of tasks both written and oral.	Checked by specific questions

2.3.2 R+ as interactive reading

Essentially R+ follows an interactive reading model in that it encourages and develops the interactions of top-down and bottom-up processes simultaneously throughout the reading process (Carrell, 1987). Top-down processes making use of skimming and scanning, prediction, thinking about the author's purpose or viewpoint and using background knowledge can be combined with bottom-up processing to further aid the decoding of a text. Bottom-up processing may be necessary for weaker readers particularly to initially recognise words prior to gaining automaticity. For stronger readers, bottom-up processing may help identify cognates, prefixes and suffixes, which may aid meaning. Bottom-up strategies may be particularly useful to Chinese L1 students learning English as L2, as in Chinese the meaning of a character is decoded before the sound, whereas in English it is sound followed by meaning. It may therefore be important to focus Chinese English learners on the sounds of English in addition to the meanings of words, which may be guided better through R+ than UR. Students with little experience of extensive reading in L2 or even in L1 may require assistance in developing effective reading strategies such that they are able to effectively employ an interactive process.

As noted by Koda (2005:5), weak readers often have inefficient word recognition skills (evaluate / evacuate), which can lead to a belief that a word is known, and therefore not noticing and acquiring, or wrongly acquiring it. In addition, Koda (1995, 1997, 1998) suggests that there may be different information-processing mechanisms involved in L2 reading by proficient L1 readers coming from different background orthographies and that these L2 readers may transfer their L1 reading strategies when they read L2. Nassaji (2003:271) suggests, therefore, that recognition of graphophonic processes should receive systematic attention in situations where the target language uses a different orthography from the readers' L1 orthography. Birch (2007:44) notes that L2 readers whose L1 does not have a Romanised alphabet, such as H.K. Chinese students, can particularly experience difficulties with orthographic and phonological correlations. This can particularly be a problem when readers move from a logographic to a sinographic system (Holm and Dodd, 1996). An effective means of developing word recognition skills may be through a programme such as R+, which could specifically target vocabulary enhancement, not only enabling more effective acquisition of vocabulary but also freeing up the readers' semantic processing, allowing them to process some syntactical items.

In a study of 60 advanced ESL students with a Farsi background, Nassaji (2003) found that lower-level component processes, such as word recognition and graphophonic processes, in addition to higher-level syntactic and semantic processes, contributed significantly to the distinction between skilled and less-skilled ESL readers. He notes that efficient lower-level word recognition processes are integral components of second language reading comprehension and that the role of these processes must not be neglected even in highly advanced ESL readers. He suggests instructional exercises that are specifically designed to target individual skills and their sub-skills, including phonological and orthographic processing skills (pg. 271). An additional suggestion by Nassaji is to combine ER with instructional opportunities (R+) where different skill areas are integrated, such that learners can have opportunities for multiple and meaningful exposures to the written text in different instructional settings.

Many Hong Kong readers who have a reliance on bottom-up strategies (Wong, 2001) may require help employing effective top-down strategies. Guided reading (R+) may more effectively encourage students to employ an interactive model when reading than UR, continuing to enhance bottom-up strategies while developing top-down processing. While some writers (Williams, 1986; Day and Bamford, 2002; Krashen, 2004a, 2004b, 2004c) have suggested that supplementation after UR is unnecessary, reading with

appropriate tasks may be able to develop an interactive reading process in an appropriate context, making both the reading and the tasks relevant to each other, rendering R+ an efficient means to develop both a reading habit and reading skills.

In addition to developing efficient word recognition skills, R+ may be employed to aid readers' background knowledge or schema, which can also aid comprehension performance (Anderson and Pearson, 1984; Carrell and Eisterhold, 1983). Laufer and Sim (1985), as noted earlier (2.1.1), stated that vocabulary was the most pressing need for L2 learners, followed by subject matter knowledge (schema), and these may influence each other.

2.3.3 R+ and schema development

Schema is used to describe the structure and role of knowledge in the mind. As with other reading models, it is a metaphor and describes a cognitive aspect of reading. Koda (2007:4) notes that successful comprehension is achieved through the integrative interaction of extracted text information and a reader's prior knowledge. According to this theory, background knowledge is vital for comprehension (Nassaji, 2007:80).

Coady (1979) states that,

...the ESL student should take advantage of his strength in order to overcome his weaknesses. For example, greater background knowledge of a particular subject could compensate somewhat for a lack of syntactic control over the language.

(Coady, 1979:11)

An interactive reading model sees the reader as an active participant in the reading process who is able to move beyond mere graphic representations of print by bringing emotions, culture, experience, knowledge and information to the reading process and not simply being involved in a psycholinguistic guessing game. If guessing is required, then the chances of accuracy may be enhanced through the use of interactive skills.

Brown suggests that the greatest problem for L2 readers is lack of content schema, and the reason why this fails, according to Carrell and Eisterhold (1983:80), is that the schema required is culturally specific and not part of a particular reader's cultural background. They illustrate this point with the line 'The runner was called out at the plate', which, unless the reader has acquired 'baseball schemata', will be largely incomprehensible.

Carrell (1983) suggests providing readers with familiar contents that include relevant cultural information; however, this could prevent them from ever developing new content knowledge (David, 2009). If readers need to activate background schema in order to better comprehend a text, it may be difficult for them to know what schema to activate until they have read the text. A number of writers (Carrell, 1985, 1988; Floyd and Carrell, 1987; David, 2009) have suggested that schema theory provides a strong rationale for the use of pre-reading activities. Williams (1987) suggests that in order to aid content schema, the teacher needs to actively build and activate it in the pre-reading, during reading and post-reading phase. This is easier to accomplish if all readers are reading the same text in a guided reading (R+) situation rather than in ER where everyone is reading a different text.

2.3.4 Tasks and the task load hypothesis

Haastrup (1991) argues that words that are more difficult to guess require greater processing effort and the more distinctive memory traces created by them leads to their being retained better in memory (2.2.2). The involvement load hypothesis proposed by Hulstijn and Laufer (2002) states that the learning and retention of unfamiliar words is dependent upon the mental effort or involvement that a task imposes. According to the theory, task-induced involvement consists of three task factors: need, search and

evaluation. The three factors may be present or absent in varying degrees but the greater the involvement-load, the better the word retention.

Need is the motivational, non-cognitive dimensions of involvement and refers to whether knowledge of a word is required for task completion. A need is moderate when it is imposed by an external agent (teacher) and strong when it is intrinsically motivated (learner needs a word for a composition and looks it up in a dictionary). Search and evaluation are the cognitive components of involvement because they entail information processing (noticing and attention). Search refers to the attempt by the learner to find the meaning of unknown words to complete a task (dictionary, teacher). Evaluation entails comparing new words with other others and making a decision with regard to its suitability to complete a task. Evaluation is moderate when learners must use a new word in a given sentence and strong when a new word must be combined with known words and used in an original context. A task's involvement load is the 'combination or the presence or absence of the factors Need, Search and Evaluation' (Hulstijn and Laufer, 2002:15). The higher the load involvement, the more effectively a word may be learned and retained. Laufer (2005) suggests that vocabulary gains are greater when learners engage in word-focused tasks compared to reading alone. The word-focused tasks may occur as the need arises. A number of studies provide empirical support for

the task factors need, search and evaluation. The studies outlined below are those applicable to reading and tasks.

When words are looked up in a dictionary (need/search – R+), recall of previously unknown words is better compared to when words' meanings are simply inferred or ignored (no search or need – ER) (Cho and Krashen, 1994; Knight, 1994; Luppescu and Day, 1993).

Using new words to write a composition (strong evaluation) results in better word learning compared to reading a text for comprehension regardless of whether new words are glossed or looked up in a dictionary (Hulstijn and Laufer, 2002; Hulstijn and Trompetter, 1998; Laufer, 2003). The implication is that *using* new words is more powerful for retention than just noticing their meanings. The studies did not take into account repeated exposures to new words that would occur naturally in an ER context in a process of noticing a new word in a variety of contexts, inferring its meaning and confirming the meaning through further exposures. Krashen (2004c:50) claims that Laufer (2003) really just compares different ways of consciously learning words and that the use of marginal glosses is unnatural.

Again, Laufer (2003, 2006) demonstrated that using new words to complete gapped sentences (moderate evaluation) and choosing the meaning of decontextualised words from several options are more effective than reading a text for comprehension. While this research suggests that tasks employing need, search and evaluation to varying degrees can lead to differential gains in word learning, the research was generally concerned with advanced learners, who may be more skilled at inferring words from context or more skilful (motivated) with tasks.

Tu (2004), in a study of six senior classes in China, compared reading comprehension, reading comprehension with filling in target words and composition writing with target words (low to high load involvement) on the retention of ten unfamiliar words. The tasks with higher load involvement resulted in greater retention, but as with other studies in incidental vocabulary acquisition (Hulstijn and Laufer, 2002; Nagy et al, 1985; Watanabe, 1997), there was a significant decline in the performance of all three groups from the immediate post-test to the delayed post-test.

While the study again confirms that incidental learning of vocabulary is possible through reading or writing tasks, they may be of limited effectiveness unless coupled with some complementary measures such as repeated exposure to target words on the

part of learners. The low scores on the delayed post-test (one week) indicate that a single exposure to a vocabulary item may not be enough to leave a sufficiently deep imprint in memory that keeps a word available for retrieval in the long term (Tu, 2004:51).

Folse (2006) suggests that the efficiency of one task (e.g. reading) relative to another is also a function of the amount of time that is required to complete it. Tu (2004) found that the initial advantage of tasks with a higher involvement load was weakened as time elapsed. While the words that were processed with a higher load involvement were retained better than words with lower involvement, all three groups showed considerable decay a week after the treatment. Tu (2004:51) suggests a ‘compelling need for measures that help consolidate the superior effect of more “involving” tasks (e.g., follow-up vocabulary exercises on the target items)’. An alternative may be a regular reading programme with tasks, as suggested by R+.

Keating (2008), in a study of seventy-nine beginning learners of Spanish, found that vocabulary retention was highest in higher loaded tasks (sentence writing) than in tasks with less load (reading comprehension); however, when time on task was considered, the benefit associated with more involving tasks faded. The task that produced the

greatest vocabulary gains, however, also suffered the greatest loss when tested two weeks after the treatment.

Folse (2006) found that the group who completed three gapped sentence activities with the same target words outperformed the group that were involved in the more involving task of writing a single sentence for each word, even when the task time was equivalent. The pedagogical implication of this study and the study by Keating (2008) is that the recycling of vocabulary is vital for long-term acquisition. Regular reading with tasks (R+) may be an effective means by which to provide the necessary repeated exposure.

Laufer (2003:273) argues against relying on reading for vocabulary growth because of the lack of available class time for the quantity of reading required to make significant vocabulary gains; however, as Krashen (2004c:50) argues, this is an argument in favour of reading, as ER can take place outside classrooms and even without teachers: 'students can continue reading long after class has finished'. The practicalities of this in the H.K. situation have already been discussed with regard to students' lack of motivation or purpose to read.

A study by Paribakht and Wesche (1997) on ESL college students was not directly an attempt to provide evidence for the load hypothesis but a comparison between reading only and reading with task-based vocabulary experiences for two hours per week over a semester. While both approaches lead to vocabulary gains, the text-based vocabulary exercises led to superior gains in L2 vocabulary acquisition. Min (2008:79) suggests that the study provides EFL teachers with an answer in how to effectively and efficiently increase ESL students' receptive vocabulary. The combining of reading with enhancement activities may be a realistic way to assist students in acquiring the basic 3,000 – 5,000 words (Paribakht and Wesche, 1997; Rott, 1999)

Min (2008) carried out a similar study to that of Paribakht and Wesche by comparing narrow reading (repeated reading of thematically related articles) with reading plus vocabulary-enhancement tasks. The subjects were fifty male Chinese speakers at a senior high school in Taiwan. The reading with tasks group demonstrated significantly more knowledge about the target language than the narrow reading group on the acquisition and retention tests even three months after the treatments ended. In this study, therefore, where the repeated reading on a theme could represent a form of ER and reading with tasks R+, there was clearly superiority for R+. Rott (1999:593) suggests that L2 learners need to be engaged in reading under both circumstances:

reading for meaning under enhanced conditions, to ensure a basic lexicon, as well as under normal conditions to advance beyond the basic requirement (Rott's research was based on German as L2).

Bell (2001) compared intensive reading with extensive reading with a group of elementary learners (no age given) in Yemen. The experimental group (n = 14) received an extensive reading programme over thirty-six hours. The control group (n = 12) received an intensive reading programme comprising short passages with a variety of comprehension and referential questions, cloze, gap-fills, multiple choice and word-building exercises. The time spent reading was carefully matched for both groups. Comprehension was measured on pre – post-test, comprising multiple choice, yes / no and cloze. The extensive reading group made three times the gains of the intensive reading group, although reliability from multiple choice and yes / no tests is not high (Meara and Buxton, 1987; Wesche and Paribakht, 1996; Waring, 1999). Bell also notes that the subjects in the ER group were aware of being involved in a separate and special reading programme, which may have influenced results. Bell suggests that intensive reading may inhibit reading improvement among learners at low proficiency levels where they focus on language manipulation rather than developing comprehension, which is genuinely developed through ER. The students spent more time on tasks than

on reading. It is important with R+ to ensure that the emphasis remains on reading and that the tasks are employed to aid reading by making text more comprehensible and enjoyable.

The importance of activities as a follow-up to reading is emphasized by both Nunan (1997:18), who suggests that they provide learners with ‘a scaffold upon which to transfer their passive vocabulary knowledge to productive use’, and Laufer and Hulstijn (2001), who stated that classroom tasks that involve learners in consistently elaborating word knowledge result in greater learning and retention than if the words are simply encountered while reading.

This section has suggested that implicit and explicit acquisition may be more effective for language acquisition than either method in isolation, and R+ is able to employ both while developing an interactive reading method that can aid both WR and SR. Pre-reading schema-building tasks may aid comprehension and employing tasks may enable greater noticing and vocabulary acquisition than from UR alone.

2.3.5 Studies involving reading with tasks

The following section reviews studies involving reading with tasks (R+): two studies that compare ER with R+ (Elley and Mangubhai 1983; Smith 2006) in a real classroom setting and a study by Lituanas et al (2001) which highlights the success of R+ for weaker readers.

The studies reviewed do not specifically involve measuring vocabulary acquisition but a variety of general language, comprehension / proficiency gains. There are few studies which do measure R+ and vocabulary acquisition directly. The purpose in reviewing these studies is to highlight the role of reading with tasks, and there is a suggestion that if a study is able to demonstrate comprehension / proficiency gains then there may also be vocabulary gains due to the high correlation between the two.

In 1979, a book-based English teaching programme was launched in primary schools on the small South Pacific island of Niue (De'Ath, 2001). The government's education system consisted of seven primary schools and one high school administered by a small Education Department. English was taught through a structural linguistic approach (TATE programme), which was almost universal in the South Pacific. The Tate Oral English Programme was introduced in Year 1, with English reading commencing in

Year 3. The scheme was rigidly controlled. The language was taught structure by structure with the pupils chanting in unison what the teacher dictated. Teaching was carried out directly from a manual to ensure accuracy.

The programme involved little natural communication or factors necessary for language acquisition and was not motivating (De'Ath, 2001). The text of the English reading books was confined to the structures previously taught in oral English lessons, was weak on plot and restricted the opportunity for students to 'meet and explore the colorful language of children's literature' (pg. 140).

The main research question was to determine which programme - the Tate Oral English with associated readers or Shared Reading - was more beneficial for Year 3 pupils (mean age 8.5 years, n = 151, six classes). The shared reading programme (Holdaway, 1979 – Appendix B2) involved students reading 48 graded readers, called 'Fiafia', written by the researcher related to island life. The books were read and discussed, prediction, recall and re-telling were encouraged and attention was drawn to vocabulary and syntax. The reading sessions were reportedly enjoyable, interactive and language rich. The stories were revisited, which increased the developmental benefits. The actual quantity of shared reading is not reported.

All students were tested on thirty-five multiple-choice sentence gap-fill comprehension questions, a fifty-word test of words drawn from the first four Tate readers and the first ten levels of the Fiafia stories and an oral language repetition test.

A simple experimental-control group contrast design was not possible, as nearly all Year 3 pupils on the island were instructed using the Fiafia stories in 1978. The language performance of Year 4 pupils was therefore assessed at the beginning of both 1978 and 1979. The 1978 cohort in Year 4 had spent the previous year learning from the Tate Program, while the 1979 cohort had learned from the Fiafia Stories.

All six classes produced substantial gains with the Shared Book method. Pupils in the new programme had improved in word recognition by 98% over the previous cohort, in oral language by 67%, and in sentence comprehension by 33%. These gains in language skills represent very strong to strong effect sizes: $r = .52$, $.40$ and $.30$ respectively.

While the study showed considerable benefits for Shared Book reading over the TATE system, one conclusion that can be drawn is that a programme that employs enjoyable communicative language activities and comprehensible input has benefits over a programme that does not.

The De'Ath study in Niue was followed by a similar large-scale study in Fiji in 1981 and 1983 by Elley and Mangubhai. The studies were fairly similar in design and the extensive 1983 programme is commented on here. Elley and Mangubhai (1983) involved 380 students in twelve rural Fijian schools in grades 4 and 5. In the control group (four classes) the students followed a traditional audio-lingual method (TATE), where they were taught to master all structures orally first. The experimental groups (eight classes) followed either a shared book approach (R+) or ER for 20 – 30 minutes each day with the remaining time on the TATE programme. Having backing for such a programme at government level allowed teachers some freedom from the rigid curriculum, such that 30 minutes per day could be deducted for reading without parent or teacher complaints - something that has been, and still is, a major problem with ER in Hong Kong. The study is significant in that it is one of the few studies to separate out the effects of silent, individually controlled extensive reading and guided reading (R+) in a real classroom situation.

At the end of the first year, the experimental groups had made significant and 'impressive' (Elley, 2000:239) gains in reading comprehension, listening comprehension, and mastery of English structures, with smaller gains in the other tests – writing and oral language. The growth in reading was twice that of the control groups.

The shared-book group performed better than the silent-reading group on a number of measures, particularly in reading comprehension and listening in Form 5.

Further testing at the end of the second year showed that the gains had increased substantially in all modes of English tested on a range of objective and open-ended type tests and the national examination results for Grade 6 pupils revealed an impressive 'spread of effect' to science, social studies, mathematics and vernacular language. The pass rate in the English examination was twice that of the control group. By the second year, there was no significant difference between the shared book group (R+) and the ER group, although they both still consistently outperformed the control group. Elley and Mangubhai (1983:66) suggest that the lack of difference between the two methods (pure ER and shared book, R+) may indicate that either the shared book activities are not as important as their advocates claim, or the teachers who used them did so ineffectively. The study reports that at times the methods overlapped, with the ER reading teachers sometimes reading aloud to their classes and the R+ teachers failing to make full use of activities. The conclusion which is drawn by the authors is that the groups that regularly read books made language gains over those that did not read but only participated in a regular class programme. The critical factors that brought about the substantial improvements were related to 'greater and repeated exposure to print in

high-interest contexts, in conditions where pupils were striving for meaning, and receiving sufficient support to achieve it regularly' (Elley and Mangubhai, 1983:66; Appendix B2). However, for all the time and training invested in the shared book method, the ER group performed just as well over the course of the study. The results suggest that students are able to acquire language from reading on their own (UR) without teacher-directed tasks.

Whilst the results are impressive, they are not particularly surprising. The control group functioned within a highly rigid, controlled L2 teaching environment. The use of interesting story books allowed the students to receive language input that was enjoyable while giving them individual responsibility and choice. As Elley and Mangubhai (1983) note, 'Theoretically.... new learning takes place at the point of interest, rather than in accordance with a carefully graded linguistic pattern' (pg. 58). The study does not resolve whether reading is better for language acquisition than other more communicative methods or language-rich environments.

While the De'Ath and Elley studies do not specifically involve a measure of vocabulary, they do show general language gains from R+, and it is suggested that gains in reading and listening comprehension, writing and oral language could reflect corresponding

gains in vocabulary (2.1.1 – reading vocabulary and proficiency / comprehension correlation).

The large-scale studies in Niue and Fiji were followed by a number of similar studies where local students were flooded with high-interest reading material employed in a shared book approach. These studies were conducted in Singapore (1985 – 1989), Brunei (1988), Sri Lanka (1995), South Africa (1997) and the Solomon Islands (1995 – 1998). In summarizing these large-scale projects, Elley (2001:238) comments that all students anywhere, regardless of their first language, age or background, can experience increased literacy through exposure to high-interest books. All of these programmes were so successful that the governments (with the exception of the Solomon islands, for which no information is available) either vastly extended them with considerable additional funding or nationalized them.

Curiously, these programmes, which involved thousands of students in dozens of classes, employed the shared book method (R+) rather than pure extensive reading (ER) even though the results from the Elley and Mangubhai (1981 and 1983) studies in Fiji had indicated that ER was just as effective as the shared book method and did not require the same amount of training. While the Elley and Mangubhai study in Fiji in the

early 1980s focused on two different experimental methods (R+ and ER), both groups had 250 high-interest readers in their classrooms. The shared book group were encouraged to read the books and thus could participate in both R+ and ER. The silent reading or ER group only had the option of reading the 250 books supplied in the classroom. They were not able to actively participate in the reading tasks on their own initiative without teacher assistance, which was not provided. The rationale for setting up large-scale studies that employed a shared reading approach rather than pure ER is possibly that a shared book programme with the addition of high interest class readers would enable students to enjoy any advantages of both methods while offering an element of teacher control and student accountability. In Singapore, the study was initiated in 30 schools in grade 1 classes. Each class received 60 books for the shared book lessons, along with guide notes on how to use them. Later in the year, 150 books for independent reading were provided 'when the pupils were ready' (Elley, 2000:240).

Elley (2000:237) notes that shared reading was an 'excellent way of ensuring that the children in a Book Flood study interacted constructively with the books every day and gradually built up their knowledge of the language, with ease and enjoyment' This is an extremely important point in favour of R+, or at least this particular type of R+ - shared reading – in that the programme *ensured* constructive interaction while gradually

developing skills and building vocabulary knowledge. Ensuring that students were exposed to language was particularly important in third-world print-deprived environments with little outside L2 exposure.

A study by Smith (2006), investigating the impact of supplementary activities on reading, showed that some supplementary activities may have a positive affect on language acquisition. The students were 15 or 16 year olds at a junior college in Taiwan. One group did extensive reading only, the second group did extensive reading with very short summaries (ER+) and the third group studied short reading passages, answered comprehension questions and analysed sentence-level components of the reading as per intensive reading (IR). As shown in Table 2.3, the ER-only group made significant gains, but only in the first semester.

The ER-only group made similar gains each semester, but the two other groups did much better in the second semester. The ER-only gain was significantly larger the first semester (compared to intensive reading (IR); $t = 2.05$, $p = .02$, one- tailed; compared to ER+, $t = 2.57$, $p = .001$), but there was no significant difference between the ER-only group and the others the second semester ($t = .838$, $t = .05$ respectively). These results

were measured on the EPER Placement/Progress Tests (Edinburgh Project of Extensive Reading (EPER - Hill 1995), which is a reading comprehension cloze test.

Table 2.4 First and second semester gains compared (Smith 2006:14)

	First semester gains	Second semester gains
IR	3.3 (7.4)	7.68 (4.9)
ER+	3.9 (7.1)	6.74 (5.2)
ER ONLY	7.56 (6.8)	6.68 (5.4)

mean and standard deviation

It is perhaps more unusual for such significant gains for ER to have occurred in the first semester (18 weeks) rather than at the end of the study, as longer-term ER programmes are generally more effective than shorter-term programmes (Krashen 2004b). All groups made similar gains in the second semester, which perhaps suggests that reading with supplementary tasks may have a part to play in enhancing students' language ability. It is unknown whether a third semester would have seen the differences between methods further evened out. The improvement of the IR group on the EPER in the second semester could be partly accounted for by the fact that the students spent their time completing tasks, such as answering comprehension questions, and analysing sentence

level components of the readings, that were somewhat related to the tasks in the test (cloze comprehension). However, this does not explain the achievements of the ER-only group.

At the beginning of the study and five months after the treatment ended, students took a local standardized language test, the CSEPT (College Students' English Proficiency Test) The CSEPT consists of sections on listening, reading, and usage and is administered annually to college students in Taiwan. Table 2.4 shows the results of the CSEPT, with a clear advantage for the ER-only group.

Table 2.5 CSEPT test results

	IR	ER +	ER ONLY
Pre	135.1 (31.8)	132.6 (32.8)	129.5 (32.9)
Post	185.8 (40.3)	181.5 (40.4)	192.8 (45.1)
Gain scores	50.7 (30.5)	49 (26.2)	63.3 (33.4)

mean and standard deviation

ER+ was the worst performing treatment, which is interesting considering that ER programmes administered in H.K. are often exactly that – reading with comprehension questions and/or summaries. This can be seen in the effect sizes, which suggest that

time spent reading only (ER compared with ER+, effect size = .54) is more effective than spending less time reading but adding related supplementary activities. (Smith does not state which effect size calculation is employed).

A study by Lituanas et al (2001) in the Philippines involved 60 students who were assigned to remedial English classes and clearly demonstrated the ability of low-ability students to acquire language from R+. The ages ranged from 12 to 18 years and ninety percent of the pupils at the school were described as coming from low-income families. The sample students were pair matched on socio-economic status, sex, IQ reading level and past achievement, with one member of each pair randomly assigned to the control group and the other to the experimental group.

The study design was a pre-test – post-test design, similar to the various Book Floods described above. Two tests used were the Informal Reading Inventory (IRI – Johnson et al, 1987), which measures reading comprehension, and a standardized oral reading test from the USA (GSORT – Gray, 1967), which measures reading speed and accuracy. The GSORT scores are associated with US class grades: thus, these students in Grade 7 would be expected to score around a 7.00 on the test, but their pre-test score was 2.45. While both tests are L1 reading tests and thus making a comparison with L1 students is

not particularly valid, their use as a pre-post test with both groups still allows for a between-group comparison that is suitable for the purposes of this study.

Both groups received forty minutes of regular daily English classes plus an additional forty-minute remedial reading class. The control group's remedial class was taught in the conventional way from a textbook, which included reading and reciting poems and focus on minimal pairs, vowel sounds, etc. They very infrequently silently read short passages from their textbook. The experimental group took part in what is described as an ER programme. In reality, the programme was more R+ than ER. Students read from a limited selection of self-selected books and then completed post-reading activities. Students spent about 45% of the time reading, with another 45% spent on activities and 10% on classroom management. The teacher spent 80% of the time helping students by answering questions and guiding students who had difficulty recognizing words. The programme ran for six months. The results showed a significant difference ($p < 0.01$) in favour of the experimental group on both post-tests. Most interesting is that the experimental group's average GSORT changed from 2.36 to 5.25 after only 43 hours of reading, which is equivalent to an improvement of almost three full school years. The reading-with-tasks programme was able to cater for a wide age range (12 – 18) of very weak students with fairly limited resources.

Lituanas et al (2001) do not discuss why they chose this particular reading method (ER+) over a pure extensive reading programme without tasks (ER), but it may have been a lack of faith in a pure ER programme for weak students. They note that,

effective ER programs seem even less common for lower achieving students, as many educators express the view that such students lack the desire and skills to read extensively. Thus, further research is needed to test the value of ER with lower-achieving students.

(Lituanas et al, 2001:217)

One can speculate that, as with Elley (2000), reading with tasks ensured that the students were actively involved in language and the programme also allowed the teacher to provide targeted guidance, as the tasks raised language questions. The students self-selected books, as with ER, and therefore each student would be working on something different, so there would be little opportunity for whole-class instruction on a task. Budget constraints may have also partially dictated the method, as whole sets of class readers are expensive and Lituanas et al note that the collection of books was assembled from a variety of donated sources and was 'barely adequate' (pg. 220).

One of the problems with studies comparing ER with non-reading control groups is that the amounts of exposure to English may not be equal for both groups. If the ER is carried out in a time-controlled classroom situation, students are always free to read in their leisure time. While from a research perspective this undermines the validity of a study, from a pedagogical perspective it is a huge advantage for ER and one of the factors in its favour in that students are able to obtain input at any time and place on their own initiative.

A study by Mason (2004) set out to show that writing as a post-reading task added little to students' language acquisition. The study employed three groups of 18 – 19 year old first year English majors at a Japanese university (n=104). The subjects all had normal English classes throughout the week and one extensive reading class per week. On average, they read about 2,300 pages over a semester from graded readers (about 250,000 words), mostly completed at home. The study is not one of pure ER, as after two weeks, listening to stories was incorporated into the ER classes. Upon completing a book, students wrote summaries of their reading either in English with correction, English without correction or in Japanese depending on which group they had self-selected. All three groups improved significantly on a 100-item cloze test (TOEIC - Test of English for International Communication), with the group that wrote summaries

in their L1 (Japanese) being the most efficient, as they spent 150 hours in total compared with the English summary group's 260 hours and the English with corrections group's 287 hours of reading, writing and rewriting for similar gains.

A large percentage of the English writing students felt that writing hindered their reading and was tiring. There was no comparison with a group that did no writing. The suggestion is that the summary writing in Japanese (L1) should not have aided the students' English (L2) acquisition, and yet these students made the same progress as the group that wrote L1 summaries. The conclusion drawn is that writing in English was of no additional benefit. Without non-reading and non-writing control groups, however, it could be argued that language gains were simply due to the students' regular classes. The actual gains from the ER/listening/writing component are not measurable. It could be argued that summary writing *per se* in whatever language was the factor improving language, and not the reading, which was unable to be measured without non-writing and non-reading control groups. What this study could suggest is a place for an L2 supplementary writing activity after reading, but its conclusion suggests that writing as a supplementary task to reading is of no benefit. Certainly in the H.K. context, with weaker readers, writing may not be effective in that it may raise the affective filter by providing an additional and more difficult task to 'pleasure' reading.

Tanaka and Stapleton (2007) set up a study in a Japanese high school of 190 first grade students (15 - 16 year olds) split into an experimental (n = 96) and a control group. The students were described as lower than average for the region on standardized tests. The experimental group participated in a home reading programme for about five to six months. The materials consisted of thirty-eight passages adapted or written by the author at the students' level. Five to ten minutes of class time was used to check some simple post-reading tasks and change books, with the reading being completed at home. The students were encouraged to read graded readers if they felt confident. There were only sixteen graded readers available to students at the start of the programme, but this had been increased to sixty titles by the conclusion of the study. Simple follow-up activities such as vocabulary quizzes and comprehension checks were included.

While the researcher was concerned that the post-reading tasks may have become a burden for the students, it was felt that with no follow-up activities, the students might not have read the materials. This is an important point in that a similar problem can occur in H.K. where students require a purpose for reading aside from pleasure or the vague notion that reading is good for them.

The topics of the reading passages were explained to the students at the beginning of each session and the passages were then read to the students.

The control group did not have the home reading but simply completed normal textbook-based tasks, which were also completed by the experimental group. Results revealed that the treatment group scored significantly higher in reading speed and comprehension than the control over a five-month period. The experimental group spent more time on English than the control group and in fact may not have needed to read the passages at all to complete the required tasks. There is no record as to the quantity of graded readers completed. The study demonstrates gains for an R+ reading programme for beginner level students, in which ER may also have played a part. The gains in reading speed and comprehension would suggest vocabulary gains although these were not specifically measured.

This section has suggested that R+ enables a learner to learn language implicitly and explicitly, benefiting not only from the quantity of input, as with ER, but also the quality. R+ may play a useful role in developing top-down and bottom-up processing and aid schema formation, which will be of benefit to both weaker and stronger readers. The task load hypothesis lends support to the value of R+ where tasks are able to

improve vocabulary learning and retention. The last portion of section 3 outlined a number of studies suggesting the effectiveness of various R+ programmes in engaging learners through tasks that are capable of developing both receptive and productive language.

2.4.0 Conclusion

This review has outlined the importance of vocabulary for L2 learners and the belief that reading extensively can aid vocabulary acquisition. It was suggested that UR in the form of ER, however, is not generally seen as particularly relevant in the H.K. education system. While there is a suggestion that weaker students in H.K. have not benefited from ER, it is not clear if this is due to a lack of motivation to read or a lack of ability to acquire language implicitly from UR. Some H.K. studies have indicated that weaker readers may not be capable of acquiring language from UR and there is a suggestion that supplemented reading may be more effective than UR for vocabulary acquisition, particularly for weaker readers.

Implicit acquisition of vocabulary from UR requires a reader to notice an unknown word, guess its meaning and retain the word for some time, during which time it is encountered again and knowledge of the word is strengthened. These assumptions are less well supported for weaker readers. Krashen's comprehension hypothesis suggests

that ER provides input, which is the key, and perhaps only necessary ingredient for language acquisition, although there are relatively few studies in support of pure ER.

There is a suggestion that quality of input may be as important as quantity.

While students may acquire some vocabulary implicitly, explicit learning may help to focus their attention on critical features, which may then be more readily acquired. It is suggested that R+ may be able to stimulate a reader's schemata of a book, which may aid comprehension, and also that the addition of tasks, which require mental effort, may lead to better and longer lasting vocabulary acquisition. R+ may be able to provide a more guaranteed method for vocabulary acquisition than the hit-and-miss approach of UR in addition to better fitting into the H.K. education system by being teacher-directed and purposeful with the furnishing of tasks. A number of studies show support for R+ over normal class programmes; however, those that compare ER with R+ generally fail to show a clear gain for R+ in spite of the additional time and effort it requires. No studies reviewed to date indicate whether UR or R+ is better suited to a particular ability level.

The questions this study seeks to answer are:

1. Can a group of weaker and stronger year 12 H.K. L2 Chinese students acquire vocabulary from uninstructed reading (UR) measured receptively after 14 days and 28 days?
2. Can the same two groups of students (WR and SR) acquire vocabulary from reading that is supplemented with vocabulary tasks (R+) measured receptively after 14 days and 28 days??
3. From which of the two methods, UR or R+, can the subjects acquire the most vocabulary - in other words, which method is more effective for vocabulary acquisition?
4. Do students prefer R+ or UR and is there any difference in their opinions between WR and SR?
5. Is there a difference between the two methods (UR, R+) for weaker and stronger students' acquisition of vocabulary? In other words, can weaker students acquire vocabulary from UR in the same quantity as stronger students and can weaker students acquire vocabulary from R+ in the same quantity as stronger students?

CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter begins by outlining the methodology of the study. The design of the quasi-experiment is then laid out with regard to the participants, materials and research procedures. The chapter concludes by considering ethical and validity issues.

3.1.0 Methodology

This study is a quasi-experiment. The study did not develop from a search of existing research in order to find a suitable ‘topic’, but rather from a real problem in a real classroom and a desire to understand and improve current practice.

The study employs a quasi-experiment, as the subjects are not a random sample but an intact H.K. secondary school class. While a quasi-experiment may have a weakness in sample selection, its strength lies in its ability to take place in real settings with real subjects and problems. The difficulty, as with true experiments, is in controlling extraneous variables such that causal claims can be made for the independent variable. This is perhaps one of the greatest challenges for quasi-experiments: whether the

population employed and the results obtained can legitimately be generalised to a broader population. While this study may have implications for the global population, its initial goal is to investigate claims pertinent to this particular group of H.K. students and their ability to acquire language from reading.

Heinsman and Shadish (1996:164), from their meta-analysis of quasi- and true experiments, found that equally well-designed studies employing either method gave comparable results. They do suggest, however, that to make causal claims based on a quasi-experimental study, the effects of initial group differences need to be taken into account. Most experiments attempt to match experimental and control groups on variables such as IQ, sex, age or study-specific skills (vocabulary knowledge). This study, however, does not employ a non-participatory control group but explores differences in the dependent variable (vocabulary acquisition) while either manipulating the independent variable method (UR and R+) or deliberately manipulating group differences (SR and WR). The study does not attempt to ensure that students in the two experimental groups are evenly matched on reading ability, but in fact ensures, through deliberate manipulation, that the two groups have significantly different reading abilities. The variance in pre-test scores between the two groups of weaker and stronger readers is controlled for through the use of an analysis of covariance (ANCOVA) where the

pre-test scores for both groups are used as a covariate (4.5.0). (An ANCOVA factors out the initial difference in pre-test scores between groups, allowing claims to be made for the dependent variable irrespective of group differences).

This study begins with the claim that language can be acquired from reading (Krashen, 1989, 2004; Elley, 2000; Day and Bamford, 1998) and that vocabulary in particular may be acquired (Nation, 2001; Krashen, 2004b; Grabe, 2009; Stahl and Nagy, 2006) and that reading with tasks may be more effective than uninstructed reading (Rosszell, 2007), particular where it employs noticing (Schmidt, 1990) and opportunities for output (Swain, 1985). It seeks first to strengthen the knowledge claim with regard to these theories for Chinese year 12 students, and then steps into relatively uncharted territory by examining whether students with a particular reading ability are better suited to a particular method.

The study seeks to discover whether students can acquire/learn a word from reading with and without tasks; therefore, it is important to define what it means to ‘learn’ a word and how we can measure its acquisition.

This study is concerned with the very beginning stages of vocabulary acquisition. Schmitt (2008) suggests that the form-meaning link is probably the first and most essential lexical aspect that must be acquired. Reading for meaning may stimulate the form-meaning link before other types of word knowledge, which may be more difficult to acquire or acquired later. This receptive knowledge of a word involves initially being able to recognise it in its written form when reading, which may also later include knowledge of its concepts, context, parts and collocations.

This study seeks to compare vocabulary acquisition from a receptive source (uninstructed reading) and a combination receptive/productive source (R+ - reading with tasks) and it was decided to measure vocabulary in its initial embryonic form-meaning state. If subjects are unable to acquire any vocabulary at this basic level, then word knowledge requiring greater understanding may be unlikely. For example, failure to attain the basic form-meaning of a word would also suggest a failure to acquire associations, collocations or any productive use of the word.

As both UR and R+ may aid receptive knowledge, as they both employ receptive acquisition, a receptive pre-test and post-test will be employed.

A test of whether a word has been learned is whether its meaning is retained over time (Waring and Takaki, 2003), not on an immediate post-test. If the reading to post testing time period is too long, it may not be possible to detect any vocabulary acquisition; external factors may also interfere with the retention of words, particularly if they are among the 3,000 most common. A time period that is too short may have few practical implications in that while it would suggest that a word has been acquired in the short term, it may well be lost before it was encountered again and able to be retained more permanently, particularly for productive use. Waring and Takaki (2003) comment that only one second language study (up to their 2003 publication date) has systematically attempted to gather data on how much learning was retained over time, so this is an area of some need, particularly as retention is seen as necessary for long-term acquisition (see also Brown et al, 2008).

This study only measures the basic form-meaning link of 26 target words. There is no measure of other words which may have been acquired, neither does the study measure the strength of knowledge of the acquired words in terms of form (spoken, written), meaning (associations, concepts) or use (grammatical functions, collocations) (Nation 2001:27). There is no measure of other gains that might be made from reading (1.3.1).

While a word is referred to as having been ‘acquired’, this only refers to the tentative form-meaning relationship that the post-test was able to detect at the time of measuring.

The study design involves a number of steps in order to compare UR with R+ within and between two groups of students (WR and SR) over two time periods (14 days and 28 days). The reader may find Table 3.12 (pg. 200) useful in providing an overview of the study and its individual steps.

The research questions are: 1. Can both WR and SR acquire vocabulary from UR measured after 14 and 28 days? 2. Can both WR and SR acquire vocabulary from R+ measured after 14 and 28 days? 3. From which method (UR and R+) can students acquire the most vocabulary? 4. Which method do the students prefer, UR or R+? 5. Is there any difference between the ability of WR and SR to acquire vocabulary from either method (UR and R+)?

The following section describes in detail the participants, materials and procedures.

3.1.1 Participants

The participants in the study are from an intact class at a H.K. secondary school taught by the researcher. The subjects are year 12 students with a further year of study to complete before attempting to enter university. They are 16 - 20 years old (mean = 17.3).

The school is a CMI school with a banding or grading of about 2.4 (Appendix B1).

The students have had approximately four hours a week of English lessons since they entered school at age five or six and currently have 275 minutes of English lessons per week. The lessons are to prepare the students for the H.K. A-level examinations, which consist of papers in listening, writing, reading and language systems (proof-reading, cloze, comprehension etc), oral and a paper that involves collating data to complete a variety of written tasks.

Prior to entering year 12, the students completed the H.K. National (year 11) English examination consisting of five papers. Students' final grades are assessed on a six-point scale from 1 to 5*, with 5* being the highest. This group of students' national examination results were: grade 4 – 3 students, grade 3 – 14 students, grade 2 -12 students and grade 1 – 1 student (n = 30). 43% of them could be classified as below average from this public English examination (grades of 1 and 2).

On an adapted version of Nation's (1990) vocabulary levels test (see test A, 3.1.2), twenty-one students ($n = 30$) had attained the most common 1,000 words (where 15 out of 18 is considered a 'pass' – Laufer and Nation, 1999), six had attained level 2 (2,000 words) and only five attained level 3 (3,000 words) with no students attaining higher levels. All students attaining a level are included in the total, thus a student may be recorded as having attained level 1 and level 2. In some instances, a student may not have attained the cut-off score for level 1 but had attained level 2.

The class of thirty students is composed of twenty-one students who arrived from other schools specifically for years 12 and 13. They were forced to leave their previous schools because their total aggregate marks were not high enough for them to stay at their higher-banded schools. One student came from an EMI school.

The students could be generally classed as non-readers. A survey of the students (Table 3.1) asked them to indicate the approximate number of books they had read in English and Chinese, excluding assigned books, over the previous year.

Table 3.1 Response table for the number of books students' read in English

0 books	1 – 2 books	3 – 5 books	6 – 10 books	11 – 15 books	16 – 20 books	20 + books

Table for students to complete

The category 1 - 2 was scored as 2 books read, 3 - 5 was scored as 4 books read, 6 - 10 scored as 8, 11 - 15 was scored as 13 and 16 - 20 was scored as 18 books read. Some students wrote numbers in the available boxes rather than ticking them (n = 29, raw data not included above in Table 3.1).

The average number of English books read over the previous year was 2.4, with no students ticking the boxes for 6 or above. For Chinese, the average was 6.7 books read over a year, with five students recording 0 books and five recording 20+. If the scores of 0 and 20+ are removed as outliers, the average number of books read in Chinese for a year is about 4 or 5. While the survey is only an approximate measure, it does suggest that the students are not reading extensively in English or in Chinese, with the exception of five students in Chinese. It is felt that students might be more likely to record a

higher number of books read than a lower number to be seen in a better light by their class teacher.

3.1.2 Materials: Placement tests

Research question 5: Is there a difference between two methods (UR, R+) for weaker and stronger students' acquisition of vocabulary? In other words, can weaker students acquire vocabulary from UR in the same quantity as stronger students and can weaker students acquire vocabulary from R+ in the same quantity as stronger students?

In order to answer this question, students need to be identified as weaker or stronger readers. This section describes the four tests employed for this purpose.

The study employed three vocabulary tests and a reading comprehension test as means to accurately describe the subjects' language ability and to group the subjects into two groups: weaker readers (WR) and stronger readers (SR).

The tests (Appendix A) consisted of a) a vocabulary levels test developed by Nation (1990) and adapted by Huang (1994, 1999) for Taiwanese students and b) a New Zealand Progress and Achievement Vocabulary Test (Darr et al, 2008), c) a self written vocabulary test (Dykes, 2008b, which tests students' knowledge of the most common

2000 words as listed by Nation (2001) and d) a New Zealand Progress and Achievement Comprehension Test (Darr et al, 2008).

The following section describes the tests and the rationale for their use.

Test A: Students' basic vocabulary level was tested on a Vocabulary Levels Test (VLT – Nation, 1990), adapted by Huang (1994, 1999) for Taiwanese students. The Vocabulary Levels Test (Nation, 1990) has been used in a number of vocabulary acquisition studies (Laufer, 1992; Horst et al, 1998; Zahar et al, 2001) and is widely accepted as a standard measure of vocabulary proficiency (Meara, 1996). The paper-and-pencil test measures receptive vocabulary knowledge at five levels of word-frequency — the 2,000, 3,000, 5,000, 10,000 and academic word level (AWL). Huang (1994, 1999) adapted the test for 9th grade Taiwanese students who had only been exposed to the most common 1,000 words in their textbooks by removing the 10,000 word level test and adding a 1,000 level test. The rationale for the change was that it was felt that Taiwanese students' vocabulary size was small and the inclusion of a test at the 1,000-word level would 'examine the subjects' knowledge of the most frequently used words better' (Huang, 1994:131). In the 1,000-word level test, 36 words were randomly selected from the 1,000 most frequent words in West's (1953) General

Service List and test items were composed following the same format as that of Nation's VLT. The adapted Huang (1994, 1999) version of the test was adopted for this study, as it was felt that a test at the 1,000-word level would provide a more sensitive measure of participants' vocabulary, as most might have failed to obtain a score at the 10,000-word level test and some might not even achieve the 2,000-word level.

The test requires students to match lexical items to their corresponding definitions and consists of six sets of six words at five vocabulary levels. Three of the words in each set are test items and three are distracters, for a total possible score of 18 at each level.

A piloting of the Huang (1994) version the previous year showed some confusion with one word-set. Students had recorded both words '4 and 5' with the meaning 'something that can help'. In the final version of the test, 'college' was changed to 'collect'

1. sing
2. sure _____ not controlled by others
3. prove _____ *something that can help*
4. college _____ without doubt
5. advantage
6. independent

The Cronbach-alpha coefficient for the test used in the current study was .859 ($n = 30$), which shows a good internal consistency (Pallant, 2007).

Table 3.2 shows the number of students passing at each vocabulary level where a ‘pass’ is 15 out of 18 words correct ($n = 30$, $M = 48.1$, $SD = 13.7$, range = 19 – 80).

Table 3.2 Passes at each vocabulary level for test A

	Level 1	Level 2	Level 3	Level 5	AWL
Total number of passes ($n = 30$)	21	6	5	0	0
Average total score	15.1	11.4	7.6	5.9	5.5

Test B: A New Zealand Progress and Achievement Vocabulary Test (PAT) is a 40-question multiple-choice (A - E) test (Darr et al, 2008). The test package includes seven different age-related tests in total, ranging from the equivalent of a native speaker year 4 to year 10. Based on the results of test A above, PAT test 3 was selected, as it provided the researcher with a test at an appropriate level for the sample students. The PAT test was chosen because the choice of seven tests meant that a test could be selected at an appropriate level. The test is for L1 speakers, but for this study there is no

attempt to compare these results with L1 students. The test is used simply to obtain a mean score for the group of subjects in order to group them as weaker or stronger readers ($n = 30$, $M = 17.8$, $SD = 5.46$, range 9 - 28).

Test C: A New Zealand Progress and Achievement Comprehension Test (Darr et al, 2008), is a 35-question multiple-choice (A – D; 2 items A - E) test consisting of seven passages and one poem. Test 3 was selected, as it provides the best coverage of years 5 – 7. The rationale for a comprehension test is that there is a positive correlation between vocabulary and comprehension (Grabe, 2009). Students will acquire vocabulary from reading and tasks that require comprehension skills, and therefore students with better comprehension skills will possibly not only have greater vocabulary knowledge, but also be more capable of acquiring vocabulary and thus be stronger readers (SR). The passages were not culturally relevant to these students, who would have little background knowledge of the passages: therefore, test scores would relate entirely to vocabulary and comprehension knowledge obtained from the passages themselves ($n = 30$, $M = 23.4$, $SD = 4.12$, range 12 - 32).

Test D: This test is a 50-question multiple-choice test (A - D) developed by Nation and Beglar (2007). The test was developed to provide a measure of learner's vocabulary size

from the 1st 1000 to the 14th 1000 word families of English. Nation and Beglar (2007:11) suggest that it is not necessary to make learners sit all fourteen levels when the test is used with elementary or intermediate learners, but that they should sit a few levels beyond their present level. The test employed made use of the first five levels, ($n = 30$, $M = 29.53$, $SD = 5.98$, range = 16 - 42).

The Cronbach-alpha coefficient for all four tests was .821 demonstrating a good internal consistency ($n = 30$, $M = 117$, $SD = 28$, range = 68 – 169, total possible = 235).

The relationship between the four reading placement tests was investigated using a Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity.

There was a strong positive correlation between all four tests, $n = 30$, $p < .01$ - Table 3.3,

Appendix A1 – A4.

.

Table 3.3 Pearson Correlation of four placement tests

Correlations					
		Levels Test Test A	PAT Vocab. Test B	PAT Comp. Test C	Nation & Beglar (2007) Test D
PAT Vocab. Test B	Pearson Correlation	.817**	1.000	.481**	.750**
	Sig. (2-tailed)	.000		.007	.000
	N	30	30	30	30
PAT Comp. Test C	Pearson Correlation	.594**	.481**	1.000	.582**
	Sig. (2-tailed)	.001	.007		.001
	N	30	30	30	30
Nation & Beglar (2007) Test D	Pearson Correlation	.837**	.750**	.582**	1.000
	Sig. (2-tailed)	.000	.000	.001	
	N	30	30	30	30
Levels Test Test A	Pearson Correlation	1.000	.817**	.594**	.837**
	Sig. (2-tailed)		.000	.001	.000
	N	30	30	30	30
** Correlation is significant at the 0.01 level (2-tailed).					

According to Cohen et al (2007) even the smallest effect size in Table 3.3, $r = .481$, is a very strong correlation (Appendix A8). The correlation indicates that all four tests are measuring the same thing: vocabulary knowledge and comprehension, which is used as an indicator of reading ability.

The above test scores were changed to percentages, totalled, and students were ranked. The bottom fourteen students made up the weaker readers' group (WR) while the top sixteen students made-up the stronger readers group (SR). The reason for the unequal groups was simply the lack of any real difference between two students where the total sample was split exactly in half.

An independent-samples t-test was conducted to compare the two groups. There was a significant difference in scores for weaker readers ($M = 94.07$, $SD = 17.58$) and stronger readers, $M = 138.5$, $SD = 16.50$; $t(28) = -7.135$, $p = .000$ (two-tailed). The magnitude of the difference in the means (mean difference = 44.429, 95% CI: -57.18 to -31.67) is very strong ($r = .80$). This means that the two groups are significantly different in reading ability as measured by the four tests.

Table 3.4 shows the scores for vocabulary test A (levels test) separated for weaker and stronger readers.

Table 3.4 Number of students passing at each vocabulary level and average scores

	Level 1	Level 2	Level 3	Level 5	AWL
Total number of WR achieving the level (n = 14)	7	1	0	0	0
WR average score	13.4	9.5	6.9	3.4	3.5
Total number of SR achieving the level (n = 16)	15	5	5	0	0
SR average score	16.6	13.5	12.7	8.1	7.3
Total achieving the level (n = 30)	21	6	5	0	0

WR = weaker readers, SR = stronger readers

The single SR who failed level 1 failed all levels but with a reasonably high average score of 47 (WR - $M = 36.7$; SR - $M = 57.3$, total $M = 47$)

3.1.3 Materials: Book Selection

Research question 1: Can a group of year 12 H.K. L2 Chinese students acquire vocabulary from uninstructed reading (UR - for a group of WR and a group of SR)?

Research question 2: Can the same groups of students (WR and SR) acquire vocabulary from reading that is supplemented with vocabulary tasks?

This section outlines the selection of the texts to be used for both UR and R+ for both WR and SR.

Using both knowledge of the students as their English teacher and the above four tests, the researcher selected a number of books that were thought to be interesting, at an appropriate level (98% of vocabulary understood) and able to generate sufficient unknown words so that acquisition could be measured through a pre-test and post-test.

Determining a student's reading level and matching a book to that level is not an exact science. Book difficulty is a combination of vocabulary, sentence structure, length, elaboration and coherence, topic familiarity and interestingness, and may be aided by

illustrations and print size (Beck and McKeown, 1989; Beck et al, 1995; McKeown et al, 1992).

The total length of a book can influence difficulty, as a less avid reader may have greater difficulty in finishing it. Particularly for students who do not read fluently, length alone can be a formidable obstacle (Grobe, 1970).

A number of readability formulas attempt to quantify text and provide an approximate school grade for a book or a score on a 'level of difficulty' scale. While some studies show that they are valid to be used in the EFL context (Greenfield, 2003), some researchers have found that classic formulas are not very accurate predictors of EFL text difficulty (Brown, 1998; Shokrpour and Gibbons, 1998).

Tests such as the Flesch-Kincaid, Fry, SMOG, Gunning-Fog and Noun Count (Elley and Croft, 1989) are only able to take into account factors related to actual words and sentences. The Fry formula, for example, is:

$$206.835 - 1.015 \times \frac{(\text{total words})}{(\text{total sentences})} - \frac{(\text{total syllables})}{(\text{total words})}$$

Such formulas take no account of how interesting a story may be to the reader. These tests may be applicable, however, to determine the rough level of a book in order to select the correct level in a series of graded readers, for example. In addition to grading the difficulty level of a text is the difficulty in accurately assessing a student's reading level and correctly matching them with a selected text. As a starting point to matching reading books with students, it was decided to examine the percentage of words in a book occurring at the 1,000 and 2,000 most common word levels. Vocabulary test A (above) had already indicated students' mastery of these words.

Short-listed books were scanned (OCR – Adobe programme), tidied and corrected where necessary, and converted to text files. Proper nouns were removed and the text files were put into Vocabprofiler (Cobb, 2008). This programme analysed the percentage of words at the 1000 and 2000 level, academic words (AWL) and off-list words (words that appear on no list – often including proper nouns). Vocabprofiler also listed all words in the text and their frequency. There was often little to differentiate between books except for their total number of words and the number of 'off-list' words, which are words that do not appear on any common vocabulary lists (proper-nouns were already removed).

Based on many years of teaching reading and experience teaching English to the age group, the researcher selected words that the students probably would not know. These were listed and books were tentatively selected based on the number of these words divided by the actual words, which gave an approximate readability level as a percentage. A number of the abovementioned readability formulas were applied, but the results were quite inconsistent with each other. The researcher ranked books on difficulty level based on his skill as a teacher, taking into account topics, particularly as they may relate to students' interests, length and perceived difficulty, and found that the readability formulas often varied dramatically from his rankings.

An analysis of one selected book, *The Promise* (Scott-Buccleuch, 2005), for example, indicated that 95.78% of the words occurred at the 1,000 or 2,000 word level. 1.42% of the words were from the academic word list (AWL, Coxhead, 2000) which was actually only eight words occurring numerous times, and none of these were included as target words, as in the researcher's experience it was thought that they would be sufficiently well known by the students. The AWL are not necessarily 'difficult' words and included words such as final, finally, goal, job, physical, professional, teams and transport. The off-list words (2.8%) were also not necessarily difficult words and

included words such as *Africa*, *bored*, *contest*, *score* and *triangle*. For this book, a total list of 48 possibly unknown words was recorded

The researcher decided on using 26 target words for each book. Waring and Takaki (2003) made use of 25 target words and Brown et al (2008) 28 target words. Some students would already know some of the 26 target words, and if the number of target words selected is too small, then there may be too few words to acquire and the post-test may not be capable of discriminating differences to a significant level. As the words would be combined for two books, too many words would make the tests too long (52 words). It was thought that at least ten unknown words would be required for each book and each student in order to significantly demonstrate gains.

In addition to ensuring a sufficient number of test words for each book, it was necessary that the selected words also appeared in similar frequencies. For example, in comparing vocabulary acquisition for UR with R+, if unknown words in an UR book all appeared ten times but the unknown words in an R+ book only appeared once, there would be an immediate advantage for UR.

35 – 55 possibly unknown words from short-listed books were presented to the students taking part in the study in a survey asking them to indicate YES, NS or NO as to whether they thought they knew the meaning of a word, were not sure or did not know the meaning (a test design employed by Horst, 2005). YES was scored as 0, NS as 1 and NO as 2. The final selection of books was based on whether there they were at a suitable level of difficulty and interest for the students according to their group (WR, SR) and whether the book had a sufficient number of unknown words for each student based on the Yes / No test. Most of the book selection was carried out in a pilot study in the preceding year, with only two unsatisfactory books from the pilot study needing to be replaced (due to low enjoyment scores – Table 3.5).

The twenty-six target words were selected as words that a large number of students had indicated that they did not or probably did not know, and target words were matched across books to ensure that words occurred at similar frequencies between books (Tables 3.8 to 3.10). In addition, there was an attempt to match the number of target words that were nouns and verbs between books, as some writers suggest that nouns are easier to learn than verbs or adjectives (Ellis and Beaton, 1993; Birch, 2007; Kweon and Kim, 2008), while others suggest verbs are easier to learn (Liu and Nation, 1985). The frequencies of words were added and divided by the total number of words in the book

to give a percentage of words known from the total text. For example, from Appendix A6, the book *Skyjack* -: the frequencies of the words added totals 236 occurrences of the 26 target words. This is 2.8% of the total 8,381 words giving a readability score of 97.2%. All books were scored in a similar way: *The Promise* – 97.95%, *Woman in Black* – 97.8%, *Eye of the Tiger* – 97.5%, *Zorro* – 97.25% and *White Fang*, 97.75%. Two words which occurred over one hundred times were not included in the count - *White Fang* (fangs) and *Zorro* (Don) as they would distort the scores and were classed as outliers.

Table 3.5 shows the final selection of books and the percentage of words within the most common 2,000 and off-list words. Upon completion of reading each book and attached to the students' final worksheet or tasks were three survey questions which asked students to mark on a 10-point Likert Scale whether they had found the book easy or difficult (easy = 1), whether they had enjoyed reading the book (enjoyed very much = 1), and how much of the book they had completed. The score for each book is also included in Table 3.5 and there appears to be little difference in readability and enjoyment between books. Appendix A6 has lists of the target words from each book and their frequencies. *Woman in Black* is read by both groups at different times employing different methods as noted in the Table.

Table 3.5 Analysis of words, difficulty and enjoyment for selected books

Book letter Code and Book title	Percentage of words at 1,000 -2,000 word level (from total words)	Total words	AWL (minus names)	Off list words (minus names)	Group, purpose and when tested	Difficulty and enjoyment (1- 10, 1 = easy/ enjoyed)
A. The Promise	95.78	8907	1.42%	2.80%	WR - UR (28 days)	4 / 3.6
B. Skyjack	95.08	8381	0.85%	3.88%	WR – UR (14 days)	3.2 / 3.5
C. Woman in Black	96.77	10908	0.11%	3.20%	WR – R+ (28 days)	4.2 / 4.7
D. Woman in Black	96.77	10908	0.11%	3.20%	SR – UR (14 days)	4.9 / 5.2
E. Zorro	95.16	12591	0.04	4.80%	All – R+ (14 days)	WR 3.9 / 4.8 SR 4.1 / 4.9
F. Eye of the Tiger	95.62	14975	0.32%	4.07%	SR – UR (28 days)	4 / 4.6
G. White Fang	93.53	12565	0.09%	6.16%	SR – R+ (28 days)	4.6 / 4.1

WR – weaker readers, SR – stronger readers, UR – uninstructed reading, R+- reading with tasks, 14 days – post-test 14 days are completing reading, 28 days – post-test 28 days are completing reading. (The Promise, Scott-Buccleuch (2005); White Fang, London (2008); The Woman in Black, Hill (2005); Skyjack, Vicary (2008); The Eye of the Tiger, Smith (2005); The Mark of Zorro, McCulley (2007).

Books for SR (D, E, F, G) were slightly longer, with a slightly higher percentage of off-list words than those for WR. Table 3.6 shows the frequency of target words and the number of nouns and verbs as target words.

Table 3.6 Number of occurrences of each target word

Name of book, group using it and purpose	Number of target words appearing once or twice only	Number of target words appearing 6 times	Number of target words appearing 7 times	Number of target words appearing 10+ times	Number target words as nouns / verbs
A. The Promise WR/UR 28	3	14	13	8	17 / 4
B. Skyjack WR/UR 14	2	12	12	8	20 / 4
Woman in Black C. WR/R+ 28 D. SR/UR 14	4	13	12	9	19 / 5
E. Zorro SR R+ 14 WR R+ 14	2	17	15	13	23 / 3
F. Eye of the Tiger SR/UR 28	3	13	11	10	23 / 0
G. White Fang SR /R+ 28	2	15	15	8	18 / 5

WR – weaker readers, SR – stronger readers, UR – unguided reading, R+- guided

reading with activities, 14 – posttest 14 days after book completion, 28 - posttest 28

days after book completion

Research question 3: From which of the two methods, UR or R+, can the subjects acquire the most vocabulary? In other words, which method is more effective for vocabulary acquisition?

To compare uninstructed reading (UR) with reading with tasks (R+), weaker readers' vocabulary acquisition performance will be compared after 14 days between books B (UR) and E (R+) and after 28 days between books A (UR) and C (R+): see Table 3.7.

For stronger readers, a comparison between methods after 14 days will be made between books D (UR) and E (R+) and after 28 days between books F (UR) and G. (R+): see Table 3.8.

Table 3.7 Word frequencies comparing UR and R+ books for weaker readers

Name of Book, group using it and purpose.	Number of target words appearing once or twice only	Number of target words appearing 6 times	Number of target words appearing 7 times	Number of target words appearing 10 + times	Number of target words as nouns / verbs
A. The Promise WR/UR 28	3	14	13	8	17 / 4
C. Woman in Black WR/R+ 28	4	13	12	9	19 / 5
B. Skyjack WR/UR 14	2	12	12	8	20 / 4
E. Zorro WR R+ 14	2	17	15	13	23 / 3

WR – weaker readers, SR – stronger readers, UR – unguided reading, R+- guided

reading with activities, 14 – posttest 14 days after book completion, 28 - posttest 28

days after book completion

Books A and C are quite evenly matched in terms of the target word frequencies of nouns and verbs between books (Table 3.7). Books B and E are not matched as well, with book E having target words occurring slightly more frequently and with more nouns.

Some of the target words in book E, *Zorro*, however, are Spanish: *hacienda*, *caballero*, *presidio*, *sombrero* and *comandante*, which have no form that the students would have encountered before. So while book E, *Zorro*, does have a greater number of target words with higher frequencies of occurrence, these words may also be a little more difficult to learn than words with which the students have some linguistic familiarity. In comparison, book B, *Skyjack*, contains words such as *handcuffs*, *binoculars*, *refuel*, *journalist* and *hijacker*. The first three words have familiar prefixes; students write journals, making *journalist* a relevant word, and *hijacker* is a familiar concept occurring not infrequently in the media. Book E has 12,591 words, while book B has only 8,381 meaning that for book E, more words need to be read to encounter each target word.

Table 3.8 Word frequencies comparing UR and R+ books for stronger readers

Name of book, group using it and purpose	Number of target words appearing once or twice only	Number of target words appearing 6 times	Number of target words appearing 7 times	Number of target words appearing 10 + times	Number of target words as nouns / verbs
D. Woman in Black SR/UR 14	4	13	12	9	19 / 5
E. Zorro SR/R+ 14	2	17	15	13	23 / 3
F. Eye of the Tiger SR/UR 28	3	13	11	10	23 / 0
G. White Fang SR /R+ 28	2	15	15	8	18 / 5

WR – weaker readers, SR – stronger readers, UR – unguided reading, R+- guided

reading with activities, 14 – posttest 14 days after book completion, 28 - posttest 28

days after book completion

There may be a slight advantage for exposure to target words for book E over book D (Table 3.8), although this is offset by book E being 2,000 words longer than D. Books F and G are quite evenly matched.

Research question 5: Is there a difference between the two methods (UR, R+) for weaker and stronger students' acquisition of vocabulary? In other words, can weaker students acquire vocabulary from UR in the same quantity as stronger students and can weaker students acquire vocabulary from R+ in the same quantity as stronger students?

To compare groups SR and WR, vocabulary acquisition will be compared after 14 days for uninstructed reading between books B and D and after 28 days between books A and F (Table 3.9).

Table 3.9 Word frequencies comparing WR with SR

Name of book, group using it and purpose	Number of target words appearing once or twice only	Number of target words appearing 6 times	Number of target words appearing 7 times	Number of target words appearing 10 + times	Number of target words as nouns / verbs
B. Skyjack WR/UR 14	2	12	12	8	20 / 4
D. Woman in Black SR/UR 14	4	13	12	9	19 / 5
A. The Promise WR/UR 28	3	14	13	8	17 / 4
F. Eye of the Tiger SR/UR 28	3	13	11	10	23 / 0

WR – weaker readers, SR – stronger readers, UR – unguided reading, R+- guided

reading with activities, 14 – posttest 14 days after book completion, 28 - posttest 28

days after book completion.

Books B and D are evenly matched. The opportunities for vocabulary acquisition may be slightly better for book F compared with book A based on word frequencies and nouns; however, this is offset by book F containing 6,000 words more than A.

To compare groups SR and WR, vocabulary acquisition will be compared after 14 days for reading with tasks between book E for both groups and after 28 days between books C and G (Table 3.10).

Table 3.10 Word frequencies comparing books between groups WR and SR for reading

with tasks

Name of book, group using it and purpose	Number of target words appearing once or twice only	Number of target words appearing 6 times	Number of target words appearing 7 times	Number of target words appearing 10 + times	Number of target words as nouns / verbs
C. Woman in Black WR/R+ 28	4	13	12	9	19 / 5
G. White Fang SR /R+ 28	2	15	15	8	18 / 5
E. Zorro SR R+ 14 WR R+ 14	2	17	15	13	23 / 3

Glossary – see previous Tables

Books C and G are well matched and book E is the same book compared for both groups, see Table 3.12). Book E, *Zorro* is used for both groups as R+. While the book will be slightly more difficult for the WR than the SR, the tasks could help to make the book more manageable for the WR.

Some of the books contained glossaries, help pages or reading tasks in the last few pages. While these tasks may have been very useful and helpful, it was undesirable for the purpose of this experiment to provide the students with this uncontrolled assistance. It was therefore decided to copy the books with these pages removed. The copied books were given to students in a plastic folder. The post-experiment briefing explained to students that permission had been obtained from the publishers for copies to be made and that copies of all books would be destroyed upon completion of the study (Appendix D).

3.1.4 Materials: worksheets and literature circle

Uninstructed reading (UR): each book had a worksheet consisting of eight simple comprehension questions whose purpose was to encourage the students to read the book in order to answer the questions and to provide a check that the books had been read. Care was taken to ensure that none of the target words were required to answer the questions. In addition to the eight questions, students were asked to complete a survey

which ascertained their opinion of the book as regarded its interest and level of difficulty (Appendix C2).

Reading with tasks (R+): Students completed two worksheets for each R+ book with one worksheet for the first half of the book and the second worksheet for the remainder of the book. The aim of the worksheet was to recycle each target word four times. Task A gave students word meanings and page numbers and they were required to find the target word to match the meaning for about half the target words from the first half of the book. Based on the Task Load Hypothesis (Laufer and Hulstijn, 2001), this task had moderate need, search and evaluation and required receptive and productive skills. Task B contained a box with the thirteen target words (or the same word family as used in the text) and sentences with missing words. Students were required to use words from the box to complete the sentences (receptive and productive) by filling in the gaps. This task required moderate need, search and evaluation. Each word would be required in three sentences. This task is based on Folse (2006), who found that vocabulary from three fill-in the-blank exercises is retained much better than one fill-in-the-blank exercise or an original-sentence-writing exercise. The second worksheet followed the same format as the first for the remaining target words for the remaining book chapters. An example from *Zorro* worksheet one is provided below (Appendix C3).

Zorro - 1

Chapters 1 – 5, pages 4 – 24

A. Can you find these words from the story?

1. Something sharp and dangerous used for fighting (page 4)
2. Used for hitting someone or something, can hurt you (page 4)
3. Hair on your top lip is called a(page 6)

This section consisted of 13 questions or half of the 26 target words.

B. Can you put a word into the sentence? You may need to use each word more than once.

moustache	sword	whip	whipping	inn	landlord	veranda
scabbard	pistol	cloak	wound	wounded	sombrero	
comandante		galloped	galloping			

1. We went to the for some food and drink. It was a comfortable place to stay.
2. He had a long on his top lip.
3. He pulled his sword out of the and ran to fight the men (see page 4).

This task made use of the same thirteen target words employed in part A. Each target word was used in three sentences.

Literature Circles - In groups of three or four, one student took the role of connector and culture collector (Appendix C1) and the others, word masters. The collector's role was aimed at building and stimulating a schema of the story, which would help to improve comprehension of the story and support the target vocabulary. The word masters were required to explain three interesting words from the story. There was no guarantee that the students would select the target words, but as they had just completed a vocabulary worksheet on the book, target words were commonly selected.

3.1.5 Materials: Pre-tests and post-tests

Vocabulary acquisition of the twenty-six words is measured by comparing vocabulary scores before reading a book (pre-test) with the scores after reading the book (post-test).

The post-tests take place either 14 or 28 days after completing the book.

The pre-test of the target vocabulary followed the same format as placement test A: thus, students would be familiar with the format. For each word, students would need to write a number to match the word with its meaning (Appendix A7). This design was similar

to that used by Tekmen and Daloğlu (2006) and Zahar et al (2001). Words that occurred more commonly in the story were mixed with less commonly occurring words to reduce the chances of guessing by elimination. Distracters of synonyms, antonyms or similar word families (e.g., butcher, grocer) were employed to make selecting the correct answer more difficult. Table 3.11 contains a sample from '*The Promise*' (WR UR28). The numbers in brackets represent the number of times the target word appears in the story (not included in the test). The distracters (items 1, 4 and 6) were chosen as words that the students probably would not know but in some cases may have seen before and recognized as a word. Words in the sample below that may be associated and thus be distracters are *wages/receipt*, *twisted/linear* and *faculty/league* (set 1). In set 2 – the 'er' ending of *beaver* could indicate a person (making clothes) or the 'fix' of *fixture* could distract the student as fixing/making clothes. In the post-test, students' answers for 'making clothes' were referee (2), magistrate (4), fixture (4) and beaver (2), with only two students obtaining the correct answer. For the meaning 'not straight', the incorrect selections were receipt (1), league (1), wages (5) and linear (1). In this example, one of the words from the story (wages) may have acted as a distracter. Guessing 'wages' incorrectly means the students cannot select that word with the correct answer (money you are paid), thereby increasing test difficulty or opportunities to guess the remaining items.

Table 3.11 Two test sets from the pre / post-test for The Promise (WR UR28)

1.	receipt	
2.	league (2)	_____ teams in a competition together
3.	wages (2)	_____ not straight
4.	linear	_____ money you are paid
5.	twisted (10)	
6.	faculty	

1.	referee (4)	
2.	magistrate	_____ making clothes
3.	fixture	_____ someone who helps you at home
4.	beaver	_____ person who starts and ends a game / game boss
5.	sew (2)	
6.	servants (4)	

Tests for the target words from two books were combined (total 52 words) and administered as a pre-test and post-test.

3.2.0 Research Procedures

3.2.1 General design

Students were split into two groups: weaker readers (n = 14) and stronger readers (n = 16) based on the four placement tests. Four books were assigned for each group based on appropriate difficulty level and interest. A group would read two books completely unguided without any assistance (UR) and two books with a variety of tasks (R+).

Twenty-six target words were chosen from each book and two books were combined into two pre-tests of 52 target words. The two post-tests were administered such that one test recorded vocabulary acquisition of the target words after 14 days and the other 28 days after the books had been collected back in. Post-tests therefore scored UR after 14 days and 28 days for both groups and R+ after 14 days and 28 days for both groups (WR and SR – see Table 3.12).

Vocabulary acquisition from the uninstructed reading is implicit, with the aim of the reading being to attract the learners' attention while minimizing disruption to comprehension. Vocabulary acquisition for reading with tasks is explicit and incidental, as while learners have their attention directed to specific target vocabulary, the subjects

are not informed of a vocabulary test and there is no necessity for them to memorize these vocabulary items, so the tasks are explicit and incidental. While the books are referred to as uninstructed (UR) and reading with tasks (R+), in fact neither method is 'instructed'. The tasks associated with R+ are to provide opportunities to notice, use and practise the target words and no formal 'instruction' is provided. Pre-reading tasks for R+ aim to stimulate and develop schema to aid comprehension. No pre-reading tasks are provided for UR.

The research procedures (steps) are noted below and summarized in Table 3.12.

1. All subjects complete three general vocabulary tests and a comprehension test to determine their general reading level and classify them as weaker ($n = 14$) or stronger ($n = 16$) readers.
2. Twenty-six words from each book are combined and tested receptively in two pre-tests of 52 items, each on different days, at least two weeks before students read the first book.

Details: The test of twenty-six target words from each book was combined into two tests of fifty-two words in the same format as that used for placement test A. Sufficient time was given for all students to complete the tests. Students moved their desks apart,

as was normal for tests, but as an additional measure to avoid any copying of answers, no two students sitting next to each other had a test where the words were in exactly the same order. Either students sitting beside each other were from different groups (WR or SR) or the words were rearranged so that, while having the same questions, the order and question numbers were different. This approach was used for all the pre and post target word tests. The pretests were carried out at least two weeks prior to step 3 (see Table 3.12). The first test was for books A, E (WR) and F, A (SR). A similar test took place about a week later for books B, C (WR), D and G (SR).

3. All subjects practised a literature circle tasks in preparation for R+. Students were given a motivational talk on language gains from reading and the importance of vocabulary.

Details: All subjects read a very short story and took part in a literature circle where they discussed the stories, following a set of discussion guidelines later to be used in the experimental stage (Appendix C1). This was not an assessed part of the experiment but was suggested from the pilot study in order to give the students some familiarity with literature circles before they began the assessed readers where they would use these discussion groups for R+. The timing of this task was not crucial, but it was carried out

after the vocabulary pre-tests (step 2), which may have helped to dissipate any residual word knowledge acquired from these tests prior to reading the first book.

Students were given a short talk, which suggested the importance of comprehensible input for language learning and the importance of having sufficient vocabulary for successful language learning. The talk explained that one enjoyable and effective means of obtaining ‘input’ and particularly vocabulary is from reading.

3.2.2 Quasi-experiment

4. All subjects independently read one of two books over a ten-day period (WR book A, SR book F). The book provided depends on their assigned reading group (WR or SR).

Research question 1: Can a group of year 12 H.K. L2 Chinese students acquire vocabulary from uninstructed reading (UR) measured 28 days after reading?

Details: Books A and F (UR) were matched as closely as possible to the subjects’ reading ability at an appropriate independent reading level, such that the SR had a slightly more difficult book than the WR (Table 3.9). Both reading books were at about

a 98% reading level of difficulty for the appropriate group, although this is not an exact science and students within a group (e.g., WR) may still vary somewhat in ability from each other. The students were given the book in class and had ten minutes of class reading time. They were instructed to read the book for homework and complete the worksheet. They were instructed that they should read for fluency, which meant faster reading where they should concentrate on the general meaning of the story and not stop to look words up. On the tenth day, the books and worksheets were collected in. The worksheet of eight questions would take not more than ten minutes for most students. After four days (free) a second book was given to students.

5. All subjects read a second book (book E) and completed pre and post reading activities (R+). The book was *the same* for all students.

Research question 2: Can a group of year 12 H.K. L2 Chinese students acquire vocabulary from reading that is supplemented with vocabulary tasks (R+) and measured receptively after 14 days? The following details relate to this research question.

Details: As book (E) was for reading with tasks (R+), it was at a slightly more difficult level than for UR. This book (*Zorro*) was the same for both groups. The rationale for all students reading the same book was primarily one of classroom management. Many teachers in Hong Kong and on the Mainland have classes of 40+ students where completing reading work in groups can be extremely difficult and is generally not practised. The book was still within the 95% - 98% difficulty level for both WR and SR, although more difficult for WR than SR. Tasks helped to make the book easier. The book was introduced with a YouTube clip where students heard the theme music and tried to guess the movie. A brief scene from the movie was shown and students recapped what they had seen. A longer clip was shown and students were provided with some background information to the story with regard to the Spanish influence and setting.

Students were given ten minutes of class time to read an initial chapter to get them started and the remainder of the book was read at home. Students completed the two vocabulary worksheets as they progressed through the book and took part in two literature circles; the first worksheet and literature circle took place halfway through the book and the second at the conclusion of the book. The worksheet answers were briefly discussed with the students once completed. Most students would have been able to complete the worksheets in about thirty minutes. The literature circle would have required approximately fifteen to twenty minutes to prepare with total group

presentation time of twenty to thirty minutes including moving chairs etc. Total time for all tasks for each R+ book would be about two hours forty minutes, with approximately one hour to one hour twenty minutes as class time (pre-reading, literature circles and worksheet answers) and the remaining time as homework over ten days (individual tasks equal to about ten minutes' work for eight school days plus reading).

6. All subjects read a third book (R+), the title depending on their groups (WR, SR), and completed pre- and post-reading activities as for step 5 above.

Research question 2: Can a group of year 12 H.K. L2 Chinese students acquire vocabulary from reading that is supplemented with vocabulary tasks (R+) measured receptively 28 days after reading? The following details relate to this research question.

Details: Book C was given to the WR and book G to the SR (Table 3.10). Both books were at an appropriate level for guided reading, which is a little more difficult than for UR. Literature circles and tasks were added to the reading for both groups as for step 5. YouTube video clips were used to introduce the settings and background of the books to each group, with some general questions to stimulate students' schema: Where is this place, do you think? What is happening? Who is the person? What are they going to do?

A set of tasks and a literature circle were completed half-way through the book and at the end of the book.

7. Subjects completed a post-vocabulary test for the first two books read. This test was for the first two books read by each group and measured UR after 28 days and R+ after 14 days. The test was the same as the pre-test for the two books. Students' desks were separated and if any students from the same group were sitting next to each other, they were given a test with the questions in a slightly different order to reduce opportunities for copying. The tests were combined and administered on the first-day of the four-day period (Table 3.12, Appendix A7).
8. All subjects independently read a fourth and final book over a ten-day period. WR read book B and SR book D. This step was carried out as for step 4 (Table 3.9).

Can a group of year 12 H.K. L2 Chinese students acquire vocabulary from uninstructed reading (UR) measured receptively 14 days after reading?

9. All subjects completed a reading survey to ascertain their opinions of the two reading methods (UR and R+). Research question 4: Do students prefer R+ or UR

and is there any difference in their opinions between WR and SR?

10. Subjects completed a post-vocabulary test for the third and fourth books completed.

This test was the same as the pre-test for the target words and was carried out as per step 7.

Table 3.12 shows the testing and reading timetable. The above design procedures (steps) are shown in the top row.

Table 3.12 Schedule for implementation of quasi-experiment

Design Steps	1	2	3	4		5		6	7	8	9	10
Number of days	17	6	4	10	4	10	4	10	4	10	14	1
Stronger Readers												
Type of reading book letter / Title			Practise Literature Circle	UR – Book F Eye of the Tiger	free	R+ - Book E Zorro	free	R+ - Book G White Fang		UR – Book D Woman in Black		
Weaker Readers												
Type of reading book letter / Title			Practise Literature Circle	UR – Book A The Promise	free	R+ - Book E Zorro	free	R+ - Book C Woman in Black		UR – Book B Skyjack		
Tests / Surveys / Talk	Vocabulary and Comprehension tests	Two Vocab. pretests	Motivational talk - 'gains from reading'						Vocabulary post-test for books A, E and F		Survey to compare attitude to UR and R+	Vocabulary post-test for books B, C, D and G

*UR – uninstructed reading *R+ = reading with tasks, Step 7 tests on day 1

The vocabulary post-test at step 7 tested UR 28 days after reading was completed (WR book A, SR book F) and R+ 14 days after reading was completed (WR and SR book E). The vocabulary post-test at step 10 tested UR 14 days after reading was completed (WR book B, SR book D) and R+ 28 days after reading was completed (WR book C, SR book G). Thus UR and R+ can both be compared 14 days and 28 days after reading for both groups. The exact number of days to complete steps 4 to 10 was adhered to strictly.

3.3.0 Analysis

Acquisition of target words is compared between pre-test and post-test within a group by means of a paired samples t-test and across a group by means of an independent samples t-test. The comparison of methods across groups employs an analysis of covariance (ANOVA). The relationship between some data is investigated using Pearson product-moment correlation. More detailed discussion of the statistical analysis can be found in Chapter 4.

3.4.0 Considerations

3.4.1 Ethical considerations

The students were not informed that they were taking part in a study, as such information would influence the results. Essentially, the students carried out normal classroom tasks (reading, discussing books and completing related worksheets), with the only addition being the vocabulary pre- and post-tests. The students were used to regular tests, both announced and occasionally unannounced. There was no non-reading comparison or 'control' group, so all students took part in all tasks. The only difference between the students was that they did not all read the same books, as the books were matched to their ability level: a normal and sound educational practice.

All results were shared with the students at the conclusion of the study with the purpose of highlighting the value of reading and the importance of vocabulary acquisition. Written permission was obtained from the students at the conclusion of the study for the researcher to publish data from the study anonymously and from the school authorities (Appendix D).

Most of the reading and worksheets were carried out in the students' own time. Some class lesson time was required for discussing answers to worksheets and for literature circles (book discussion). The students' normal examinations involve reading, listening, writing and oral components. The tasks the students were involved in should have aided rather than hindered their development in these areas.

3.4.2 Considerations of validity

There is a single class of students completing either the same or similar tasks at the same time depending on whether they are designated as stronger or weaker readers. The groups (SR/WR and UR/R+), in effect, act as their own control, meaning that internal validity is high: thus, there can be confidence that changes in the dependent variable (vocabulary test scores) are due entirely to the independent variable (reading methods), rather than the influence of extraneous variables. Environmental factors and differences in intervention are all controlled for through this design. A single class of students controlled by the experimenter gives confidence to the external validity or the degree to which generalisability can be made from a particular experimental condition to other populations or settings (Cohen et al, 2007). Carefully detailed explanations of the exact procedures and accurately described statistical procedures allow for replication on

different populations, strengthening the knowledge claim and allowing for generalisability.

The length of time spent on R+ and UR is not equal: therefore, one could expect greater vocabulary gains for R+ and this needs to be considered in the final analysis. No controls were placed upon dictionary or word list use, although students were encouraged to read the uninstructed books for pleasure and not to use a dictionary. The students may have consulted a dictionary at any time and this was not recorded.

The length of time taken to read a book could not be controlled. While students had ten days to complete each book, a student may have completed it at any time within this period, which would mean that the post-test 14 days after returning the book may have in fact been considerably longer for some students. However, claims that vocabulary can be acquired and retained are still valid. Few of the students are avid readers and some may have struggled to complete the book within the ten-day period.

CHAPTER 4

RESULTS AND ANALYSIS

4.0 Introduction

This section analyses data for the five research questions. There are a considerable number of calculations, and to make it easier for the reader, these have been put into a table format where possible. All assumptions applicable to the statistical analysis have been met unless otherwise stated. The data was analysed using SPSS 16.

Standard recording of results includes the reporting of statistical significance or testing of the Null Hypothesis. Schmidt and Hunter (2002:65) state that ‘significance testing almost invariably retards the search for knowledge by producing false conclusions about research literature.’ The reason for this, Cohen (1990) suggests, is that the Null Hypothesis is never true for social science data because two random samples will nearly always have slightly different means. With a large enough sample, a mean difference of .00001 could be statistically significant. Valentine and Cooper (2003) note that significance tells us very little about the practical significance or relative impact of one variable upon another. The significance of a result depends on the sample size of the observed data (Field, 2009). A large enough sample may indicate that a mean difference of .00001 is significant ($p < .01$) but be unable to detect significance in a large mean

difference of a small sample. Significance figures are also rather arbitrary with 'importance' attributed to an effect of $p < .04$ and $p < .06$ considered an unimportant effect, although with equal sample sizes, the effects are likely to be very similar (Field, 2009). For this reason, effect sizes are calculated for this study for all calculations where there seems a credible need.

An effect size provides the strength of a relationship between variables. This study makes use of Pearson's correlation coefficient ' r '. Field (2009:332) suggests that it is widely understood and frequently used. An ' r ' effect size will lie between 0 (no effect) and 1 (a perfect effect). If $r = .10$, this is a small effect, as the effect explains just 1% of the difference between the means, while $r = .30$ is a medium / strong effect (9% of the total variance) and $r = .50$ is considered a large effect (25% of the variance – Cohen, 1992). For readers more familiar with Cohen's d , Appendix A8 has a conversion formula and relationship table. Effect size is a measure of the effectiveness of the treatment. Thompson (2002) argues against simplistic interpretation of effect size at cut-off points (small, medium, large) and suggests that there is more value in relating then to prior studies and confidence intervals. Confidence intervals are provided for all calculations but this study has few peers with which the effect sizes can be compared.

Field (2009) defines a confidence interval (CI) as the range of values around a given statistic that is believed to contain the true value of that statistic. The 'true value' means not just the sample statistic (M) but the value you would get if, theoretically, you could calculate it for the entire world. A 95% CI means that there is a probability of 95% that the CI contains the true mean. If, for example, our mean difference between two samples is 10.54 (see 4.2.1) and 95% CI ranges from 13.40 to 7.68, we can be 95% confident that the true value (total population) and our calculated mean difference are similar. Put simply, Field (2009:48) states that a confidence interval (CI) for the mean is a range of scores constructed such that the population mean will fall within this range in 95% of samples. (Two negative signs can be made positive in a CI as they occur due to the way the figures are placed into the formula). If a CI includes zero (CI -1.4 to 3.5) then we are 95% confident that the true mean may in fact be zero which means no difference between the means

The codes SR14 and WR28 refer to the group of readers (stronger or weaker) and the number of days after reading that the post-test was administered (14 or 28 days). Following the level of significance ($p < .05$), the significance level has been given (.000) to allow the reader to see the relationship, if any, between the significance and the effect size. All t-tests are two-tailed. UR refers to uninstructed reading and R+ is reading with

tasks. SRUR14, therefore, would refer to stronger readers using uninstructed reading tested after 14 days.

4.1.0 Results analysis (RA) for research question 1

Can a group of weaker and stronger year 12 H.K. L2 Chinese students acquire vocabulary from uninstructed reading (UR) measured receptively after 14 and 28 days?

The effective acquisition of vocabulary was measured on a receptive test after 14 days and 28 days. A paired-sample t-test (dependent-means) was used to determine whether there were significant differences between the pretest and the post-test, or in other words, whether students had acquired a significant quantity of vocabulary from uninstructed reading.

4.1.1 RA: weaker readers after 14 days

A paired-samples t-test was conducted to evaluate the impact of reading on the acquisition of target vocabulary words. There was a statistically significant increase in vocabulary scores from the pretest ($M = 6.43$, $SD = 2.70$) to the post-test ($M = 10.14$, $SD = 2.88$), $t(13) = -5.091$, $p < .0005$ (.000 – two-tailed). The mean increase in scores

was 3.71 with a 95% confidence interval (CI) ranging from -2.14 to -5.29. The effect size ($r = .82$) is very strong.

4.1.2 RA: weaker readers after 28 days

A paired-samples t-test show no significant vocabulary gains for weaker readers (WR28) from the pretest ($M = 11.14$, $SD = 3.80$) to the post-test ($M = 12.43$, $SD = 4.50$), $t(13) = -1.84$, $p > .05$ (.089 – two tailed). The mean gain in vocabulary was 1.29 with a 95% CI of -2.80 to .22. The effect size ($r = .45$) is very strong. This was the first book students read, and three readers from each group (WR and SR) reported that they had read less than 50% of the book. Even though the students had read less than 50% of the book (15 chapters in total), 21 out of 26 target words (81%) appeared at least once in the first seven chapters. However, these words were repeated in later chapters and the repetition may have been helpful for acquisition. The raw data indicated that the ‘non-readers’ had not acquired any vocabulary and in some cases their post-test scores were lower than their pretest scores. A further paired-samples t-test with the scores for these three students removed showed a significant difference for WR28 from the pretest ($M = 11.18$, $SD = 4.17$) to the post-test ($M = 13.36$, $SD = 4.52$), $t(10) = -3.32$, $p < .05$ (.008). The mean increase in scores was 2.18 with a 95% CI ranging from -3.65 to -.716. The effect size ($r = .86$) is very strong.

4.1.3 RA: stronger readers after 14 days

For stronger readers, a paired-samples t-test showed significant vocabulary gains (SR14) from the pretest ($M = 13.19$, $SD = 3.67$) to the post-test ($M = 16.69$, $SD = 3.07$), $t(15) = -6.14$, $p < .0005$ (.000). The mean increase in scores was 3.50 with a 95% CI ranging from 2.29 to 4.72. The effect size ($r = .86$) is very strong.

4.1.4 RA: stronger readers after 28 days

A paired-samples t-test showed no significant vocabulary gains for stronger readers (SR28) from the pretest ($M = 11.25$, $SD = 4.09$) to the post-test ($M = 11.44$, $SD = 4.15$), $t(15) = -.267$, $p > .05$ (.793). The mean increase in scores was .19 with a 95% CI ranging from -1.69 to -1.31. The effect size ($r = .07$) is weak.

As noted above for WR, three students read less than 50% of the book. Only 13 out of the 26 target words (50%) appeared in the first 8 chapters (total 16 chapters) so again the three non-readers were removed. When these three scores were removed, there was still no significant difference from the pretest ($M = 11.23$, $SD = 4.4$) to the post-test ($M = 12.15$, $SD = 3.96$), $t(12) = -1.528$, $p > .05$ (.152). The mean increase in scores was .92 with a 95% CI ranging from -2.239 to .393. The effect size ($r = .40$) is strong. The fact that the CI contains zero suggests that the samples are equal.

A group of year 12 H.K. L2 Chinese students can acquire significant vocabulary from uninstructed reading (UR) measured on a receptive test. Both WR and SR showed significant gains from uninstructed reading after 14 days with very strong effect sizes. After the removal of three readers from each group who had read less than 50% of the books, both groups also demonstrated gains after 28 days with strong to very strong effect sizes. The CI for SR however indicates the mean difference may include zero.

Brown et al (2008:138) report the acquisition gain from a total of 'available words'. This means, for example, that WR14 who had a pretest score of 6.43 out of a possible 26 had 19.6 available words to learn. The percentage of words acquired from total available words is included in Table 4.1, which also shows the overall vocabulary gain for UR for both groups on both tests.

Table 4.1 Vocabulary gains from Uninstructed Reading

Group	Period of time (days)	N	Pre	Sd	Post	Sd	Gain	% gain from 26 words	% gain from total available words
WR	14	14	6.43	2.70	10.14	2.88	3.71	14.3	18.95
WR	28	11	11.18	4.17	13.36	4.52	2.18	8.3	14.7
SR	14	16	13.19	3.67	16.69	3.07	3.5	13.5	27.32
SR	28	13	11.23	4.4	12.15	3.96	0.92	3.5	6.2

Students' raw marks revealed that some students had failed to acquire any vocabulary from UR. For weaker readers, two students' scores remand stagnant after 14 days; however, both of these students did show gains on the book measured after 28 days, so they were capable of acquiring some language from UR.

One stronger reader failed to make any gains after 14 days. This student had the highest total score in the placement test (20), which would possibly have made gains more difficult with a ceiling effect. Fourteen students' post-test scores were either the same as their pretest or lower after 28 days. Six of these students (three WR and three SR) admitted that they had read less than half of the book and their scores were removed

from all calculations related to UR measured after 28 days, as noted above. With these scores removed, one WR and two SR failed to acquire any vocabulary at all from the two UR books when students' total gains were added together. Table 4.2 shows the number of students in a group who had failed to acquire vocabulary when tested 14 or 28 days after uninstructed reading.

Table 4.2 Number of students failing to acquire vocabulary from uninstructed reading

Group and test delay (days)	Number of students failing to acquire vocabulary <i>(same or negative post-test scores)</i>
WR 14	2
SR 14	1
WR 28	3 + 3 non-readers
SR 28	7 + 3 non-readers

After reading two books (UR14 and 28), students on average gained 2.5 words with a range of -1 – 10. The student with the lowest overall total scores from the four placement tests only acquired four words in total from the two unguided books, although another student with a better vocabulary knowledge only acquired one word

and the student with the best vocabulary knowledge in the class acquired no words from UR. This may have been partly a ceiling effect, as with a score of 20 in the pretests, the student only had six available words for acquisition. The relationship between vocabulary knowledge and acquisition of vocabulary from reading is discussed further in 4.4.0.

Some students were more capable than others of acquiring vocabulary from UR but there appeared to be no difference between WR and SR.

4.2.0 Results analysis (RA) for research question 2

Can a group of weaker and stronger year 12 H.K. L2 Chinese students acquire vocabulary from reading that is supplemented with vocabulary tasks (R+) and measured receptively after 14 and 28 days?

The effective acquisition of vocabulary was measured after 14 days and 28 days. A paired-samples t-test was used to determine whether there were significant differences between the pretest and the post-test, or in other words, whether students had acquired a significant quantity of vocabulary from reading with tasks (R+).

4.2.1 RA: weaker readers after 14 days (R+)

For weaker readers after 14 days, a paired-samples t-test showed significant vocabulary gains from the pretest ($M = 6.50$, $SD = 3.10$) to the post-test ($M = 16.64$, $SD = 6.25$), $t(13) = -7.93$, $p < .0005$ (.000). The mean increase in scores was 10.54 with a 95% CI ranging from -13.40 to -7.68. The effect size ($r = .91$) is very strong.

4.2.2 RA: weaker readers after 28 days (R+)

A paired-samples t-test after 28 days showed significant vocabulary gains for weaker readers from the pretest ($M = 10.50$, $SD = 3.32$) to the post-test ($M = 17.43$, $SD = 4.48$), $t(13) = -6.07$, $p < .0005$ (.000). The mean increase in scores was 6.92 with a 95% CI ranging from -9.61 to -4.24. The effect size ($r = .86$) is very strong.

4.2.3 RA: stronger readers after 14 days (R+)

For stronger readers after 14 days, a paired-samples t-test showed significant vocabulary gains from the pretest ($M = 10.62$, $SD = 3.67$) to the post-test ($M = 22.81$, $SD = 3.54$), $t(15) = -11.65$, $p < .0005$ (.000). The mean increase in scores was 11.31 with a 95% CI ranging from -13.74 to -8.87. The effect size ($r = .95$) is very strong.

4.2.4 RA: stronger readers after 28 days (R+)

A paired-sample t-test after 28 days (R+) showed significant vocabulary gains for stronger readers from the pretest ($M = 7.75$ $SD = 3.28$) to the post-test ($M = 16.38$, $SD = 4.0$), $t(15) = -8.36$, $p < .0005$ (.000). The mean increase in scores was 8.63 with a 95% CI ranging from -10.83 to -6.43. The effect size ($r = .91$) is very strong.

4.2.5 RA: all students after 14 days (R+)

All students read *Zorro* (SR 14 and WR 14) as an R+ book. Pretest and post-test scores for both groups were combined and a paired-sample t-test showed significant vocabulary gains for all students from the pretest ($M = 8.70$ $SD = 3.96$) to the post-test ($M = 19.93$, $SD = 5.81$), $t(29) = -8.36$, $p < .0005$ (.000). The mean increase in scores was 11.23 with a 95% CI ranging from -9.55 to -13.63. The effect size ($r = .93$) is very strong.

4.2.6 Summary and comparison tables for research questions 1 and 2

A group of year 12 H.K. L2 Chinese students can acquire significant vocabulary from reading with tasks (R+). Both WR and SR showed significant gains from R+ after 14 and 28 days. Effect sizes for all measures are very strong. Table 4.3 shows the overall vocabulary gain for R+ for both groups. As with Table 4.1 (UR), Table 4.3 also shows the percentage of words acquired from total available words for R+.

Table 4.3 Vocabulary gains for Reading with Tasks

Group	Period of time (days)	N	Pre Test	Sd	Post Test	Sd	Gain	% gain	% gain from total available words
WR	14	14	6.50	3.10	16.64	6.25	10.14	39.0	52
WR	28	14	10.50	3.32	17.43	4.48	6.93	26.6	44.7
SR	14	16	10.62	3.67	22.81	3.54	12.91	46.8	83.9
SR	28	16	7.75	3.28	16.38	4.0	8.63	33.1	44.3
All students	14	30	8.70	3.96	19.93	5.81	11.23	43.2	64.9

Table 4.4 summarizes the effect sizes for research questions 1 (Table 4.1) and 2 (Table 4.3). Strong to very strong effects are evident for all groups and methods. There appears to be little difference in the strength of the effects between methods or groups, and this is analysed in research question 3 (4.3.0).

Table 4.4 Effect sizes for research questions 1 and 2.

	Uninstructed Reading (UR)		Reading with Tasks (R+)	
	Weaker Readers	Stronger Readers	Weaker Readers	Stronger Readers
14 days	.82	.85	.91	.95
28 days	.72	.40	.86	.91

effect sizes: Appendix A8

Table 4.5 shows the number of students in a group failing to acquire vocabulary after the 14-day or 28-day test after reading with tasks.

Table 4.5 Failure to acquire vocabulary from R+

Group and test delay	Number of students failing to acquire vocabulary (<i>same or negative post-test scores</i>)
WR 14	1
SR 14	0
WR 28	1
SR 28	0

After reading two books (R+), students on average gained 18.9 words (WR = 17 words, SR = 20.8 words) with a range of 2 – 34 words. The student with a gain of only two words had the lowest aggregate score from the four placement tests and scored 0 at the 1,000 common word level. This student acquired four words from UR28 and two words from R+28. The subject's total gain was six words from four books.

The relationship between subjects' total scores in the four reading placement tests and the acquisition of vocabulary from reading the four books was investigated using a Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity.

There was no correlation between vocabulary level and acquisition from the books, $n = 30$, $p > .05$ ($p = .180$). This suggests that students with better vocabulary knowledge did not necessarily acquire more vocabulary from reading than those with less initial vocabulary knowledge.

Table 4.6 combines results for research questions 1 and 2 and shows total vocabulary gains for UR and R+ for both tests. Table 4.7 shows the same results ranked according to their percentage gains. Students acquired at least three times as many words for R+ than UR.

Table 4.6 Comparison of UR with R+

Group	Period of time	Gain UR	% gain UR	Gain R+	% gain R+
WR	14	3.71	14.3	10.14	39.0
WR	28	2.18	8.3	6.93	26.6
SR	14	3.50	13.4	12.91	46.8
SR	28	0.92	3.5	8.63	33.1

From all four books, students gained on average 24 words ($SD = 2.90$) with a range of 6 – 37.

Table 4.7 Listed in ascending order of percentage gain from total words available

Group	Method	Period of time	% gain	% gain from total available words
SR	UR	28	3.5	6.2
WR	UR	28	8.3	14.7
WR	UR	14	14.3	18.95
SR	UR	14	13.5	27.32
SR	R+	28	33.1	44.3
WR	R+	28	26.6	44.7
WR	R+	14	39.0	52
SR	R+	14	46.8	83.9

Both weaker and stronger students were able to acquire vocabulary from UR and R+.

The gains for R+ for both groups appear to be greater than for UR for both groups and

this is dealt with further in research question 3 (4.3.0). The total gains for WR R+ after 28 days (44.7%) were greater than the gains for SR UR after 14 days (27.32%). This is noted in Chapter 5.3. WR appear to have acquired more vocabulary from UR than SR. SR appear to have acquired more vocabulary than WR from R+. This is discussed in detail in research question 5 (4.5.0).

4.3.0 RA for research question 3

From which of the two methods, UR or R+, can the subjects (WR and SR) acquire the most vocabulary? In other words, which method is more effective for vocabulary acquisition UR or R+?

A one-way between-groups (method) analysis of variance (ANOVA) was conducted to explore the difference between UR and R+ for each group. In the weaker readers' group, the difference in vocabulary acquisition between methods was statistically significant ($F_{3, 49} = 11.681, p < 0.001$). In the stronger readers' group, vocabulary acquisition was also significant ($F_{3, 57} = 33.101, p < 0.001$). Post-hoc comparisons using the Tukey test were carried out to determine where the differences lay. As for RA 4.2.2 and RA 4.2.4, the three 'non-readers' for both WR and SR were removed for UR28.

4.3.1 RA: weaker readers R+ and UR after 14 days

The Tukey test indicated that for weaker readers using R+, the mean score after 14 days ($M = 10.14$, $SD = 4.79$) was significantly different from UR14 ($M = 3.71$, $SD = 2.73$), $p < .005$ (.000), with a very strong effect size, $r = .68$.

4.3.2 RA: weaker readers R+ and UR after 28 days

For weaker readers after 28 days, R+ ($M = 6.93$, $SD = 4.27$) was significantly different from UR 28 ($M = 2.18$, $SD = 2.18$) $p < .05$ (.001) with a very strong effect size, $r = .61$

4.3.3 RA: stronger readers R+ and UR after 14 days

The Tukey test indicated that for stronger readers using R+14, the mean score ($M = 12.19$, $SD = 4.18$) was significantly different from UR14 ($M = 3.50$, $SD = 2.28$), $p < .001$ (.000), $r = .69$, which is a very strong effect size.

4.3.4 RA: stronger readers' R+ and UR after 28 days

For stronger readers on R+ 28, the mean score ($M = 8.63$, $SD = 4.13$) was significantly different from UR28 ($M = 0.92$, $SD = 2.18$) $p < .0005$ (.000) with a very strong effect size, $r = .79$.

4.3.5 Group comparisons

Table 4.8 shows a summary of the results for the ANOVA post-hoc tests comparing UR with R+ (RA 4.4.0 compared with 4.4.4). The table shows a gain for R+ over UR for both groups with all effect sizes considered very strong.

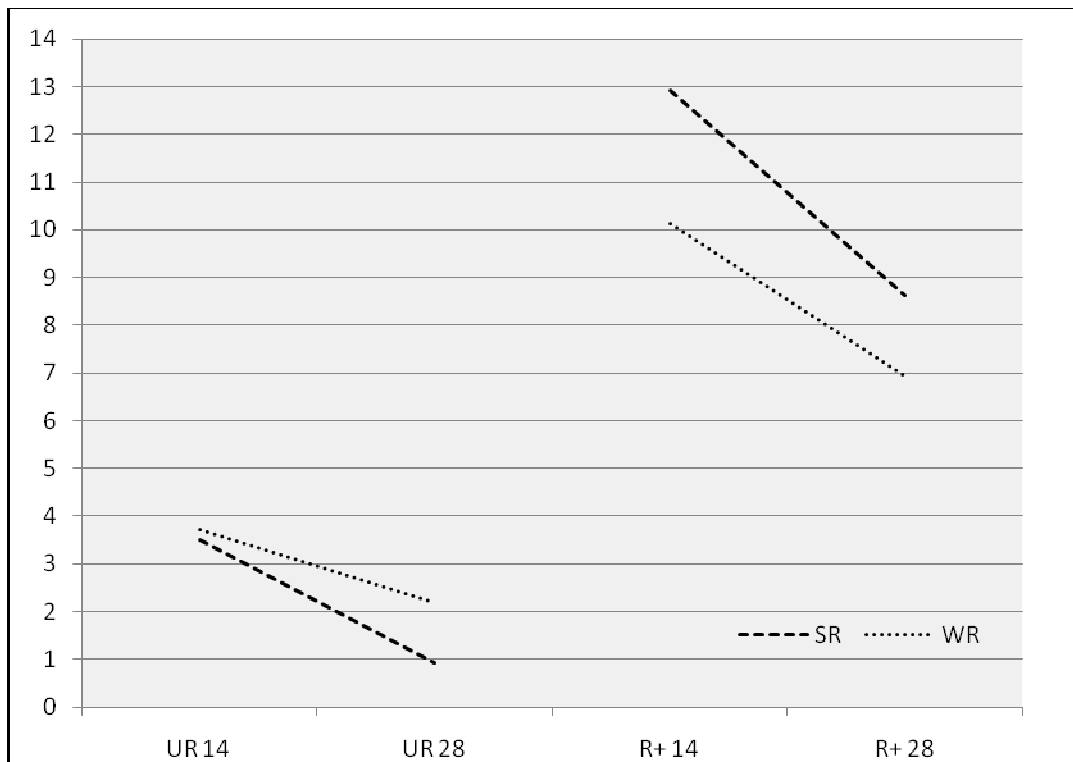
Table 4.8 ANOVA post-hoc test effect size summary comparing UR with R+

Group	Test days	Advantage for	Effect size ' <i>r</i> '
WR	28	R+	.61
WR	14	R+	.68
SR	14	R+	.69
SR	28	R+	.79

effect sizes: Appendix A8

Figure 4.1 shows vocabulary acquisition for both groups and both methods. Vocabulary decay can be observed between 14 days and 28 days for both UR and R+. Superior overall vocabulary acquisition can be seen for R+. The lines suggest a difference in vocabulary decay for UR 14 to UR 28 between SR and WR, showing a greater decline for SR than for WR.

Fig. 4.1 Vocabulary acquisition for SR and WR for UR and R+ measured after 14 and 28 days



4.3.6 RA: vocabulary decay for UR

An independent samples t-test was carried out to compare the mean differences between the two groups for vocabulary decay for UR. The scores for the non-readers UR28 were removed, as were their corresponding scores for UR14. There was no significant difference in scores for vocabulary decay for SR ($M = -2.69$, $SD = 2.81$) and WR ($M = -1.00$, $SD = 3.49$); $t(22) = 1.32$ $p = .202$ (two-tailed). The magnitude of the differences in the means (mean difference = 1.69, 95% CI: -.975 to 4.36) was moderate ($r = .27$). The CI includes zero which means that there is a 95% chance that our mean difference

may be zero so we can not say with any confidence that there is any difference in vocabulary decay between the two groups for UR.

The slope of the lines (fig.4.1) appears to be slightly steeper for R+ than UR, suggesting a greater loss of vocabulary. For SR on R+, the line appears steeper than for WR on R+.

A comparison between methods is carried out in 4.3.7 and 4.3.8 and an analysis comparing decay for R+ between SR and WR is carried out in 4.3.9.

4.3.7 RA: vocabulary decay across methods (WR)

An independent samples t-test was carried out to compare the mean difference of vocabulary decay between the two methods. For weaker readers, there was no significant difference in scores for vocabulary decay for UR, $M = 1.00$, $SD = 3.49$ and R+, $M = 3.21$, $SD = 6.84$; $t(23) = -.975$, $p = .340$ (two-tailed). The magnitude of the differences in the means (mean difference = 2.21, 95% CI: -6.91 to 2.48) is moderate ($r = .20$). This suggests that the WR lost slightly more vocabulary from R+ than from UR, although they had gained more vocabulary. The CI includes zero which means that there is a 95% chance that our mean difference may be zero so we can not say with any confidence that there is any difference in vocabulary decay between the two methods.

4.3.8 RA: vocabulary decay across methods (SR)

For stronger readers, there was no significant difference in scores for vocabulary decay for UR ($M = 2.54$, $SD = 2.79$) and R+, $M = 3.56$, $SD = 5.84$; $t(22) = -.620$, $p = .542$ (two-tailed). The magnitude of the differences in the means (mean difference = -1.02, 95% CI: -4.45 to 2.40) is modest ($r = .12$). SR lost a little more vocabulary from R+ than UR, but the effect is not statistically significant and the CI suggests no difference.

4.3.9 RA: vocabulary decay across between SR and WR on R+

An independent samples t-test was carried out to compare the mean difference of vocabulary decay between groups on R+. For R+ there was no significant difference in scores for vocabulary decay for WR ($M = 3.21$, $SD = 6.84$) and SR, $M = 3.56$, $SD = 5.84$; $t(28) = -.150$, $p = .882$ (two-tailed). The magnitude of the differences in the means (mean difference = .348, 95% CI: -5.09 to 4.39) is small ($r = .02$). This suggests that there is no difference in vocabulary decay for R+ between groups.

While slightly more vocabulary was lost from R+ than from UR, R+ registered greater initial gains, suggesting that losses are greater where gains are greatest. There was no significant difference between methods or between groups for vocabulary decay.

4.4.0 RA for research question 4

Do students prefer R+ or UR and is there any difference in their opinions between WR and SR?

The questions are reproduced below from Appendix A9 for easy reference. Each question was scored on a Likert scale (1 = strongly agree, 10 = strongly disagree).

Questions 10 and 11 helped to profile the students as ‘readers’ or ‘non-readers’ and were discussed in 3.1.1. Question 12 was analysed qualitatively and the results are noted below.

1. I prefer reading books on my own without any help, worksheets or discussion.
2. I think that reading books on my own without help can improve my English.
3. I prefer reading books with some discussion and exercises afterwards to help my English.
4. Reading a book and doing work after it will help my English better than just reading the book.
5. Just reading books for fun with no work afterwards is useful.
6. I don't have time to read English books.
7. I don't think reading books will really help me in my study.
8. I will read a book even if there are no grades or assessment.

9. I will only read books if I have to do it.
10. How many English books did you read for fun last year in form 5? *Don't count the books you HAD to read as part of your class work*
11. How many Chinese books did you read for fun last year in form 5? *Don't count the books you HAD to read as part of your class work*
12. If you don't read English books, what are your reasons?
- a)
- b)
- c).....
- More.....

Likert scores were tallied for each survey question and scores for WR were compared with SR by means of an independent-samples t-test (n = SR:15, WR:14), which found differences between WR and SR on questions 5 and 9.

Table 4.9 Survey questions with tallied Likert scores

Question	1.	2.	3.	4.	5.	6.	7.	8.	9.
WR	5.4	4.5	3.6	2.9	5.7	5.6	7.2	3.8	5.5
SR	4.7	5.1	3.9	3.5	4.3	5.4	7.9	3.2	6.7
Average	5.0	4.8	3.8	3.2	5.0	5.5	7.5	3.5	6.1

4.4.1 RA for survey questions

Scores for questions 1 and 2 are clustered around 5.0 and suggest that students are not really sure whether they preferred R+ or UR or whether one is better than the other for improving their English. Questions 3 and 4, however, indicate a preference for R+ and a belief by many that it can improve their English better than UR. There appears to be little separating the opinions for WR and SR on these four questions.

An independent-samples t-test for question 5 showed no significant difference between WR ($M = 5.69$, $SD = 1.55$) and SR ($M = 4.27$, $SD = 2.40$), $t(26) = 1.83$ $p > .05$ (sig: .078). The magnitude of the differences in the means (mean difference = 1.43, 95% CI: -.174 to 3.026) is medium / strong ($r = .34$). There is a suggestion that stronger readers were more inclined to accept that reading for pleasure (without tasks) was more useful than were weaker readers, who indicated that they thought tasks were more

necessary. The CI includes zero however, which means that the results are inconclusive.

The results for questions 2 and 3 in Table 4.9, suggest that WR showed a slightly greater tendency towards a preference for tasks after reading.

Students generally accepted that reading would help them with their study (Q.7, $SD = 1.86$) and that they would read without extrinsic rewards (Q.8) or being forced to read (Q.9). The results may be partly influenced by the briefing at the beginning of the study, which outlined the advantages of reading for language benefits. However, while students generally indicated that they thought reading was beneficial and would do it independently, question 10 revealed that on average, they had only read 2.4 books in English in the previous year (see 3.1.1). An independent-samples t-test for question 9 showed no significant difference between WR ($M = 5.54$, $SD = 2.03$) and SR ($M = 6.67$, $SD = 2.26$), $t(26) = 1.38$, $p > .05$ two-tailed, (sig: .177). The magnitude of the differences in the means (mean difference = 1.13, 95% CI: -2.805 to .549) is moderate / strong ($r = .26$). While this suggests that WR may be a little less inclined to read without some form of coercion than SR, the CI includes zero which suggests that the mean difference may be zero.

4.4.2 RA for survey question 12: Reasons for not reading

For Q.12, two students wrote that they wanted to be forced to read, which perhaps suggests that they believed that reading was useful but were unable to motivate themselves or find the time for reading. During the pilot study (n = 30) debriefing session, a student also mentioned that he thought reading was useful but that he would not do it unless forced.

Students' answers to question 12 were collated and grouped under headings in Table

4.10.

Table 4.10 Reasons for not reading

Reason	1.	2.	3.	4.	5.	6.	7.	8.	9.
	No time	Not Interested	Too lazy	Hate English	Too difficult	Prefer other activities	Boring	Don't know what to read	No use
WR n = 14	10	3	2	2	5	3	3	1	0
SR n = 15	7	2	1	0	2	4	6	1	1
Total (n= 29)	17	4	3	2	7	7	9	2	1
%	59	14	10	7	24	24	31	7	3

The preference for other activities included computer games, Chinese books / newspapers for weaker readers and Internet reading, songs and non-fiction (2) for stronger readers. The reasons for not reading could primarily be summarized as lack of

time (59%) and lack of interest (not interested, hate English, 21%) in addition to lack of suitable material (too difficult, boring, 55%). While question 6 indicated that 'time' was not a problem for students, this result showed that it was a major factor in not reading. Only 10% of answers for WR indicated boredom as a factor in not reading, compared with 27% for SR.

4.5.0 RA for research question 5

Is there a difference between the two methods (UR, R+) for weaker and stronger students' acquisition of vocabulary, measured receptively after 14 and 28 days? In other words, can weaker students acquire vocabulary from UR in the same quantity as stronger students and can weaker students acquire vocabulary from R+ in the same quantity as stronger students measured after 14 and 28 days?

The WR and SR are significantly different in vocabulary and reading comprehension as measured on the four preliminary placement tests. An analysis of covariance (ANCOVA) allows the differences between the groups to be explored while statistically controlling for the influence of group difference on the dependent variable. Pallant (2007) recommends the use of ANCOVA with a two-group pretest/post-test design with the scores on the pretest treated as the covariate to control for pre-existing differences

between the groups. Pallant (2007:291) suggests that it is useful in situations where there are quite small sample sizes and subjects have not been randomly assigned to groups. Mason and Krashen (1997) and Yamazaki (1996) employed an ANCOVA in order to control for differences in pretest scores between their control and experimental groups in a study of vocabulary acquisition from UR, as did Lee (2007a: 154) to factor out differences between groups.

A one-way between-groups analysis of covariance was conducted to compare vocabulary acquisition between stronger readers and weaker readers on R+ and UR. The independent variable was group (WR, SR) and the dependent variables were post-test scores on R+ and UR after 14 and 28 days. Participants' pretest scores were used as the covariate in the analysis. Scores for UR 28 were removed for the three students from each group who had read less than 50% of the books.

Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variance and homogeneity of regression. After adjusting for pretest scores, the following results were calculated:

4.5.1 RA: WR and SR compared on uninstructed reading (14 days)

For UR 14, vocabulary acquisition is equal for weaker and stronger readers and there is no significant difference ($F_{1,27}=3.813$, $p > 0.05$, (significance = .061), $r = .35$). This is a medium - strong effect size, suggesting that weaker readers attained more vocabulary from uninstructed reading (UR) measured after 14 days than did stronger readers (see figure 4.1). Table 4.1, however, indicates that the difference in terms of raw scores is very small (.21 words) and that in terms of percentage gains from total available words, SR made the greater gains. With the ANCOVA factoring out pretest scores, $r = .35$ is still worth noting.

4.5.2 RA: WR and SR compared on uninstructed reading (28 days)

For UR 28, vocabulary acquisition is equal for weaker and stronger readers and there is no significant difference ($F_{1,27}=1.191$, $p > 0.05$). However this was the first book students read, and as noted above, three students read less than 50% of the book. When these three scores were removed, ($F_{1,21}=2.043$, $p > 0.05$, (sig .168), $r = .29$ which is a medium effect in favour of the weaker readers.

4.5.3 RA: WR and SR compared on reading with tasks (14 days)

For R+ 14, vocabulary acquisition is equal for weaker and stronger readers and there is no significant difference ($F_{1, 27} = 2.889$, $p > 0.05$), significance = .101, $r = .31$, which is a medium - strong effect, suggesting that stronger readers attained more vocabulary from reading with tasks measured after 14 days than WR. Table 4.3 shows a larger percentage gain from total words available for SR. Both groups read the same book and their perceived enjoyment and difficulty scores for the book were almost the same (Table 3.5, book E).

4.5.4 RA - WR and SR compared on reading with tasks (28 days)

For R+ 28, vocabulary acquisition is equal for weaker and stronger readers and there is no significant difference ($F_{1, 27} = 0.052$, $p > 0.05$, (sig .821), $r = .042$, which is a very small effect.

Table 4.11 ANCOVA pre and post-test mean scores

Method	R+				UR			
Test period	14		28		14		28	
	pre	post	pre	post	pre	post	pre	post
Weaker Readers	6.5	16.64	10.5	17.43	6.43	10.14	11.18	13.36
Stronger Readers	10.63	22.81	7.75	16.38	13.19	16.69	11.25	12.15

4.6.0 Summary of results

Both weaker and stronger students were able to acquire a significant number of words from uninstructed reading (UR) measured after 14 days. WR were also able to show significant gains with a very strong effect size after 28 days. WR were able to acquire more vocabulary than SR from UR. While the difference is not significant, it is a moderately strong effect after 14 days and a medium effect after 28 days. The study was unable to detect vocabulary decay for either group from 14 days to 28 days although the raw data indicated that SR lost slightly more vocabulary (ten words) than WR (six

words). Three readers failed to make any gains from UR14 and sixteen from UR28, which was the first book read, and this number included six 'non-readers'.

Both weaker and stronger students were able to acquire significant quantities of vocabulary measured after 14 and 28 from R+. While there was vocabulary decay from 14 to 28 days, both groups (WR and SR) declined at a similar rate. Only two students failed to acquire any vocabulary from R+.

Students acquired significantly more vocabulary from R+ than from UR, although a little more time was spent on R+. Students acquired as many as 2.5 to 9 times as many words from R+ as from UR. There was no correlation between a student's vocabulary level and their ability to acquire vocabulary from reading.

While the data suggests that weaker readers were more inclined towards favouring R+ than stronger readers, with stronger readers believing that UR was more useful than R+, the results are inconclusive. Similarly, there is a suggestion that WR enjoyed UR more than SR but again, the results are inconclusive. In spite of the students' belief that reading without tasks is useful, very few of them read in English for pleasure.

The primary reasons given for not reading English books were a lack of time, lack of interest and motivation and the lack of suitable materials.

While it can be stated that there was no significant difference in vocabulary decay between methods and groups or between the methods favoured by a group, the results are inconclusive. The lack of significance (p) can be contributed, in part, to a small sample size. While the effect sizes may suggest a positive correlation between the dependent and independent variable, if the CI includes zero there is a 95% chance that there is no difference between the means. Caution needs to be exercised in stating that there is no significant difference as the lack of statistical power (due to the small sample size) can increase the chance of a Type II error. This is when there is a difference between the means but it is unable to be detected.

CHAPTER 5

DISCUSSION

5.0 Introduction

This section explains and discusses the results from Chapter 4 for each research question in turn. Where possible, the results are compared and contrasted with similar studies. Limitations of the study are also discussed.

5.1.0 Research question 1

Can a group of H.K. students acquire vocabulary from uninstructed reading?

The study by Tekmen and Daloğlu (2006) of three groups of Turkish L2 university students (2.1.3), described as intermediate, upper intermediate and advanced, recorded vocabulary gains of 10.4%, 13.3% and 17.7% respectively on a similarly designed post-test administered immediately after reading compared with 14.3% (WR) and 13.5% (SR) for the current study. The students were considerably more advanced than the students in the current study, with 51% of their lower group having attained Nation's 2,000-word level, compared with 31% of the current study's stronger readers (Table 3.4) and 7% of the weaker readers (Nation, 2001).

The scores in the Tekmen and Daloğlu study may be slightly inflated, as the testing was completed immediately after reading and the students firstly listened to the target book being read to them while they followed along and then had time to read it again themselves. Brown et al (2008) showed an advantage for listening while reading over reading only. Listening to the story guaranteed that all students had been exposed to the target words at least once and possibly twice after reading the book again themselves. In the current study, students simply self-reported whether they had completed a book or not, and there were no guarantees as to the accuracy of their reporting. The text used by Tekmen and Daloğlu was only 2,500 words long, so there was a greater opportunity to encounter the thirty target words in a short time and repetitions of target words would be condensed and encountered more rapidly compared with the current study where the reading of a 9,000 word text was spread over ten days with twenty-six targeted words tested (WR).

Factors reducing the scores on the Tekmen and Daloğlu study were the inclusion of thirty distracters in the test mixed with the thirty target words making guessing more difficult. Tekmen and Daloğlu (2006) employed thirty target words, but the pre-test mean score for the lower group was 15 (50%), meaning that they could only acquire 15 words ($SD = 4.2$). Pre-test scores for the current study, by way of contrast, were 6.4

(27%) for the weaker group, allowing the students a chance to acquire more words (20 available words out of 26).

Brown et al (2008) suggest analysing lexical gains by calculating the number of words acquired from the total number available (total number of words minus pretest score = total number of words available: see 4.1.4 and Table 4.1). When the scores are analysed in this way, Tekmen and Daloğlu record gains from total words available of 20.8% and 28.3% on an immediate post-test for their two lower groups and 48.5% for the advanced group, although this group may have suffered from a ceiling effect. The current study recorded gains from total words available of 19% (WR) and 27% (SR) after 14 days. Brown et al (2008), however, report very little vocabulary decay on a one-week delay and a three-month delay post-test. The findings in the current study are therefore in keeping with the findings of Tekmen and Daloğlu for unguided reading.

Tekmen and Daloğlu (2006:232) noted that the highest group learned significantly more words from the reading passage than did the two lower groups. This was not the case in the current study, where WR acquired slightly more vocabulary than SR after 14 and 28 days although gains from total available words were higher for SR after 14 days but not after 28 days (Table 4.1). Even though Tekmen and Daloğlu's study showed that the

ability levels of the three groups of students varied significantly, they all read the same text. A fundamental premise to acquiring language from reading is that the material must be at an appropriate level of difficulty, which is about 95% – 98% comprehensible (Schmitt, 2008; Liu and Nation, 1985) but closer to 98% (Laufer and Ravenhorst-Kalovski, 2010), or as Krashen (1981) suggests, $i + 1$. While every effort was made in the study to ensure that the book read by the students (*The Golden Fleece*) was appropriate for all of them, it is quite likely that it was either too difficult or too easy for some students.

Horst et al (1998) also noted that capable subjects with larger L2 vocabulary sizes had greater incidental word learning gains than lower ability subjects who were unable to build their lexicons sufficiently through incidental vocabulary acquisition (2.1.3). Horst et al, like Tekmen and Daloğlu (2006), also made use of a single text for a group of 34 low-level students and they note themselves that ‘some in the group must have found it challenging’ (pg. 213).

In both Tekmen and Daloğlu (2006) and Horst et al (1998), the students read the book in a single sitting, which may have aided acquisition by allowing immediate cumulative acquisition of new words. In the current study, by comparison (UR28), the stronger

students had only 14.8 available words to acquire from an 11,000-word novel and the weaker students 14.7 words from an 8,900-word novel read over ten days (Table 4.1). Where students read a book over a number of days, a new word may be encountered on day one and then not encountered again until several days later, depending on how frequently and for how long the student reads. Spreading the relatively small number of words over ten days may have reduced the intensity of repetition compared with studies where all the words were met in just an hour or two from a smaller total number of words. Therefore, acquisition comparison between shorter intensive reading with an immediate post-test and longer texts read over a longer period may not be particularly valid. Extended reading may make it more difficult to acquire words. However, the current study attempted to replicate real life, where students would more than likely read a whole book over several days.

In the current study, the text was matched as accurately as possible to students' ability, demonstrating that lower ability subjects are able to build their lexicons sufficiently through incidental vocabulary acquisition as well as, or better than, students with more ability providing the text is at an appropriate level of difficulty. An additional explanation for the success of the WR may relate to how the books were perceived by the students in terms of difficulty and enjoyment.

An analysis of the two books used for UR14 for the WR and SR (B and D) in Table 3.9 shows the books as being very similar with regard to frequency of target words and nouns. There was a difference in how the students perceived the books in terms of readability and enjoyment (Table 3.5), neither of which was measured by Tekmen and Daloğlu (2006) or Waring and Takaki (2003). Weaker readers' (UR14 book B - *Skyjack*) average score for enjoyment is 3.5 (1 = enjoyed very much) and for perceived difficulty 3.2 (1 = very easy). For stronger readers (UR14 Book D – *The Woman in Black*) the scores are 4.9 and 5.2 respectively (Table 3.5).

5.1.1 Comparison of UR14 book difficulty and enjoyment

An independent-samples t-test was conducted to compare the difficulty levels between UR14 books for SR and WR. There was no significant difference between WR book B (*Skyjack*) $M = 3.23$, $SD = 1.09$ and SR book D (*The Woman in Black*) $M = 4.90$, $SD = 2.33$, $t(21) = 2.287$, $p = .058$ (two-tailed). The magnitude of the difference in the means (mean difference = 1.67, 95% CI: -.066 to 3.42) was a very strong effect, ($r = .52$). While the effect size suggests that the WR found their book easier than the SR, the CI includes zero meaning that we can not detect a difference between the difficulty of the two books with any certainty. It is suggested that the size of the effect justifies further research (with a larger sample) to estimate group differences more precisely (4.6.0).

An independent-samples t-test for enjoyment showed no significant difference between book B (WR – *Skyjack*) $M = 3.46$, $SD = 1.61$) and SR book D (*The Woman in Black*) ($M = 5.29$, $SD = 2.61$), $t(21) = 1.848$ $p = .063$. The magnitude of the difference in the means (mean difference = 1.73, 95% CI: -.101 to 3.58) was strong, ($r = .40$). The CI is quite wide and includes zero meaning that we can not detect a difference between the difficulty of the two books with any certainty, as with 5.1.1.

5.1.2 Comparison of book enjoyment and difficulty for UR28

For the first books read, UR 28 (book A and F), there is little difference in the frequency of the target words (Table 3.9) although there is a slight advantage for the number of nouns for *Eye of the Tiger* (SR); however, WR reading *The Promise* acquired more vocabulary (Table 4.1). In terms of difficulty, both books scored 4, but *The Promise* (WR) was perceived as more enjoyable with a score of 3.6 (1 is ‘very enjoyable’) compared with *The Eye of the Tiger* (SR), which scored 4.6 (Table 3.5).

An independent-samples t-test was conducted to compare enjoyment for UR28 books for both groups. There was no significant difference between book A (WR – *The Promise*) $M = 3.64$, $SD = 1.12$) and book F (SR - *Eye of the Tiger*) $M = 4.56$, $SD = 2.40$, $t(18) = 1.32$, $p = .313$. The magnitude of the difference in the means (mean difference

= .92, 95% CI:-.998 to 2.837) was strong ($r = .31$). The CI is quite wide and includes zero meaning that we can not detect a difference between the difficulty of the two books with any certainty, as with 5.1.1. The books were scored the same for difficulty.

WR were able to acquire vocabulary from UR at a similar rate to SR after 14 days and slightly better after 28 days ($r = .29$). There is a suggestion that perceived enjoyment / difficulty of the text may have made a contribution towards vocabulary acquisition, although this claim is fairly tentatively made from a limited sample and further research is needed.

The claim for the current study is that vocabulary acquisition is possible from unguided reading where the text is matched to student ability and interest and the test allows for gains to be demonstrated. The vocabulary gains are only detected at the very basic form-meaning level.

5.1.3 Assumptions of UR

The first two assumptions for the claim that language can be acquired from UR (2.2.2) argue that the learner must notice words (Schmidt, 1990) and be able to guess the

meanings of them correctly (Laufer, 2003). The results indicate that a number of previously unknown words were correctly guessed.

While students were able to notice and correctly guess the meanings of some words, most of the target words were not correctly guessed. For WR14, from 19.6 available words, 15.9 words (81%) were not guessed correctly, and for SR14, from 12.8 available words, 9.3 (73%) were not correctly guessed. For the first WR novel of 8900 words, students only acquired 3.7 words on average. It is likely that more than just four words were noticed, as Wong (2001) suggests that H.K. students feel that they need to know the meaning of every single word when they read. However, if all the unknown target words were noticed, they were not acquired. The conclusion is that noticing on its own may not be sufficient for acquisition.

It could be argued that grouping all the target words together from one book into a test may have provided students with a context, which aided their ability to guess the correct answers. The following questions are from *The Promise* (WR UR 28) pretest / post-test.

1. hospice
 2. opponent _____ junior level, does something for fun not money
 3. amateur _____ a bed to carry someone on when sick
 4. ailing _____ the other team
 5. stretcher
 6. nursery
-

1. miserable
 2. hum _____ not happy
 3. amiable _____ loud, high sound made by blowing
 4. whistle _____ place where sport is played
 5. stadium
 6. situate
-

It might have been possible for a student to realize that they had seen some of these words before and identify them (opponent, amateur, stretcher, miserable, whistle, stadium), which would imply an element of noticing. However, even if a student did recognise all of the target words, they would still need to be correctly matched with their meanings. The 26 target words for this book were combined with 26 target words

for a different book (R+14), which was tested at the same time, which may have made associating the target words with a particular book a little more difficult. While Folse (2004) suggests that partial or completely incorrect inferences may be made, this test design may have enabled students to reject previously incorrect guesses (which were not provided) and to more accurately guess from the choices given.

Some words occurred more frequently in the book than others and may have been more readily recognized. In the above example, 'whistle' appeared six times in the story and 'stadium' eleven times. While a reader may have been able to more readily recognize these words, they still needed to find their meanings in the test and guessing the remaining words would still have been difficult with four words to choose from and the answer 'miserable' only appeared twice in the story. The word 'opponent' occurred five times in the book and while it may have been more easily matched, it would not have helped the student to guess the remaining answers in the set ('amateur' and 'stretcher'), which only occurred once in the story. We would conclude that noticing unknown words occurred from reading and the significance of the paired-samples t-test for UR would suggest that readers had attained some clues as to the basic form – meaning of the word to be able to guess meanings correctly from those provided.

The uninstructed texts combined were rated by the students with an average difficulty level of 3.9 (1 = very easy, 10 = very difficult) so students should have been able to read them reasonably quickly for general meaning, as they were instructed to do without resorting to dictionaries. Students may well have skipped unknown words (Birch, 2007); however, the test results suggest that this was not the case for all 26 targeted unknown words, as some were acquired.

Rott (1999) indicated that only two encounters with unfamiliar words during reading significantly affected learners' vocabulary growth (in German) but six exposures produced significantly more vocabulary knowledge. However, other writers suggest that learners need to meet an unknown word as many as ten to fifteen times before it is learned and this depends on a variety of factors (Nagy et al, 1985; Nagy et al, 1987; Shu et al, 1995; Nation, 2001:237). Both UR books for SR had thirteen target words appearing six or more times, with eight or nine words appearing ten times and WR books had twelve or more words appearing six times and eight appearing ten times or more. Based on frequency alone, there was an opportunity for both groups to acquire significantly more words than they did. Tekmen and Daloğlu (2006), however, noted that frequency of occurrence was less of a factor in word acquisition for their lower ability group than for the middle ability group.

To test the relationship between word frequency and acquisition, a Pearson product-moment correlation for WR UR14 between word frequency and number of correct answers was carried out. Preliminary analyses were performed to ensure that there was no violation of the assumptions of normality, linearity and homoscedasticity. There was a strong positive correlation between the two variables, $r = .57$, $n = 14$, $p < .005$. This means that there is a correlation between the words acquired and their frequency of occurrence for weaker readers after 14 days. After 28 days, the correlation was not significant, and although weaker, $r = .35$, $n = 14$, $p > .05$, it is still a medium effect. For the SR there was no correlation between word frequency and acquisition for either book (SR14: $r = .14$, $n = 16$, $p > .05$, SR28: $r = .19$, $n = 16$, $p > .05$). This is the opposite of what Tekmen and Daloğlu (2006) concluded. Frequency of occurrence was more of a factor in word acquisition for the lower ability group for UR.

While words with higher frequency in the text were more likely to be acquired by WR this was not guaranteed. For example the word ‘colonel’ occurred twenty-nine times but only four students correctly guessed it in the post-test. The word ‘cabin’ occurred fourteen times for WR14, with two students acquiring it. A greater number of occurrences helped weaker readers to acquire words more than the stronger readers. SR

may be better able to discriminate which words are important, whereas WR are more influenced by frequency of occurrence alone.

The third assumption to be accepted in support of vocabulary acquisition from UR is the ability of the reader to retain the vocabulary that they have firstly noticed and secondly correctly guessed the meaning of. WR readers demonstrated a gain of 18.95% from total available words 14 days after reading and 14.7% after 28 days or decay of 4.25% or 1.5 words. For stronger readers the gain was 27.3 % from total available words 14 days after reading and 6.2% respectively a loss after 28 days of 21% or 2.6 words). Brown et al (2008) recorded a total loss of 12.46% after one week and 11.37% after three months. The studies were quite different in their designs and population samples, so comparisons are difficult, particularly as the current study measured decay from day 14 to day 28. Both studies do show that some newly acquired words were able to be retained for at least one month.

This is important, as it is unknown how long it would take before the newly acquired words could be recycled in another book and reinforced. The longer words are able to be retained, the greater the chance of a reader encountering them again. Weaker readers would have a greater chance of sighting new words again simply because the number of

different words is fewer and they occur more frequently (1,000 – 3,000 most common).

Students may have acquired partial knowledge of some words that the test did not record knowledge of. The current study only measured 26 target words and there may be other words which were acquired or which had their tentative knowledge strengthened. The strength of word knowledge was only recorded on a receptive word - form type of test. Possibly the words retained after 28 days may have been better 'known' in terms of their form, meaning and use and may have been able to be used productively, although this study did not measure the strength of word knowledge.

The first book students read was uninstructed and tested after 28 days. Students had ten minutes of class time to start reading the book. Students rated the first books (A and F, Table 3.3) as a 4.0 for difficulty and 3.6 – 4.6 for enjoyment (1 being the easiest and most enjoyable). In spite of the motivational talk (Table 3.12 – step 3) and despite the fact that the majority of students found the book relatively easy and enjoyable, six students (20%) read less than 50% of the book and all of these students showed no vocabulary gain from the pretest to the post-test. The students who failed to read failed to acquire any vocabulary. While this may be an obvious conclusion, it is nevertheless a vital point. Lai (1993b) concluded from her H.K. study that the weakest students read the least and also made the least gains. In the current study, the 'non-readers' were

classified as both WR and SR, although compared with Lai's groups they would have mostly been weaker students.

5.1.4 Motivation to read

Perhaps the greatest problem with UR, particularly in the form of ER, is actually getting the students to read. Waring says of his Japanese university students:

If we don't require them to read and they aren't going to do it on their own, then how can we start to instill a love of reading?' However, even if we 'require' students to read it can be very difficult to enforce it and monitor it. If we make reading part of a course requirement then it can still be difficult to ensure that it has been completed.

(Waring, 2010)

If all students are reading different books, as in an ER programme, then a formal post-reading test or task can be difficult to administer and monitor. While book reports are commonly used to check that reading has been completed, they are easily completed without the need to necessarily read the book, as noted by Mason and Krashen (2004). Lai (1991) suggested that the reason why a high-level reading group in her H.K. study

had failed to make gains was that they were in the habit of skimming through books quickly in order to get enough information to complete the post-reading task and obtain credit for having ‘read’ the book. The tasks themselves can assume greater importance than the targeted reading where they are used to monitor or assess reading.

Table 4.10 is reproduced below from Chapter 4 as Table 5.1 with shading added to indicate the possible reasons why some students may not have read the first book in the study (UR28).

Table 5.1 Reasons for not reading with possible reasons for not reading the *first book*
highlighted

Reason	1. No time	2. Not Interested	3. Too lazy	4. Hate English	5. Too difficult	6. Prefer other activities	7. Boring	8. Don't know what to read	9. No use
WR n = 14	10	3	2	2	5	3	3	1	0
SR n = 15	7	2	1	0	2	4	6	1	1
Total (n= 29)	17	4	3	2	7	7	9	2	1
%	59	14	10	7	24	24	31	7	3

The students found the first book they read (books A and F) to be reasonably easy (both rated 4) and reasonably enjoyable (rated 3.6 and 4.6) so lack of interest (2) too difficult (5), boring (7) and don't know what to read (8) would seem to not be relevant here. After the motivational talk, it would be surprising if students immediately thought reading was of no use (9). It is suggested that the reasons for not reading the first book relate primarily to attitudes toward reading, such as lack of time (1), too lazy (3) and a preference for other activities (6). The fourth and final book students read in this study was also uninstructed, although all students reported that they had completed it even though SR reported that it was slightly more difficult than the first book they had read, albeit a little more enjoyable. Possibly a more precise and direct means of determining why the students did not read the first book would have been to ask them directly; however, it is only when various parts of the study are collated much later that some information comes to light, by which time it is a little too late to get reliable information.

The students were not very familiar with reading an English novel for pleasure. Initially some (six non-readers, 33%) appear to have had a lack of time or desire for reading in English, but with the provision of interesting books and an element of compulsion required by the study, participation increased, with all students completing the final UR

book. While the results confirm that the students who did not read failed to acquire vocabulary, they also suggest that some students who reported that they had read the books also failed to acquire vocabulary (Table 4.1), meaning that vocabulary acquisition may not be guaranteed from UR. Further research could investigate why some students fail to acquire vocabulary from some books.

5.1.5 Comprehensible reading level

The total number of words the researcher thought the students would not know from the WR28UR book (*The Promise*) was initially forty-eight. Thirty-three of these were tested further in the preliminary Yes/No test with twenty-six target words finally selected. However, if the students did not know forty-eight words in total and acquired 14.7% of them after 28 days (as in the current study) then the total acquisition from the book may have been seven words (it was 2.2 for the study). If students retained these words, they would need to read 142 novels of a similar length (9,000 words) to acquire a thousand words, taking almost four years. For WR14UR (*Skyjack*) there was a total of thirty-six possibly unknown words. With the study demonstrating a gain of 18.95% (Table 4.1) from total available words the students may have also acquired seven words from this book.

The outcome of the calculations from this study is in contrast to the theory of a number of writers (Nation, 2001; Stahl 1999; Schmitt, 2000; Stahl and Nagy, 2006) who suggest that if students know 98% of the total words (9,000) they would be exposed to 180 new words in the book, and if they learned 1 in 10 they would acquire 18 words from the book. In fact, students acquired less than a quarter of this number (WR UR14, 3.71 words acquired). Stronger readers were able to acquire 27.32% of the total available words, which would mean they would be capable of acquiring forty-nine words from the book based on this calculation. The researchers suggest an acquisition of 10% from the total unknown words, whereas after 14 days on average for both groups over four books (14 + 28 days combined), the students acquired 17%. The estimate that seems incorrect is that students should know 98% of the words for a book to be at the 'right level' (Hu and Nation, 2000).

The Promise (Book A) has 8,907 words: if students know 98% of the total words, there would be about 270 unknown words as opposed to the 48 the researcher originally listed. While *The Promise* only had about 48 unknown words, the students rated it as 4 for difficulty (1 is easy). It is felt that a rating below 5 is appropriate for uninstructed reading. The other UR books used in this study (books B, D and F) had from 35 – 60 possibly unknown words (excluding proper nouns) and were also rated below 5 for

difficulty. If a book had 180 unknown words, the students would probably have rated it well above 5 for difficulty, which may not be acceptable for UR.

The Promise, with about 9,000 words, has about 36 pages with about 250 words per page. If students do not know 2% of the words, this is 180 words or five unknown words per page. An old maxim from the researcher's primary school teaching days was the five-finger rule. Hold up a finger for each word on a page you don't know. Five fingers up and the book is too difficult. The maxim is still in use today (Nettles, 2006:291; Mosenthal et al, 2001:23; Preator, 2010; O'Brien, 2011) with one school (Anon, 2009) setting book levels as:

0-1 Fingers - The book is too EASY

2-3 Fingers - The book is at the Interest level.

4 Fingers - The book is at the Challenge level. You can try it ~ be sure it makes sense.

5 Fingers - The book is at the Frustration level and is not a good choice for now.

These calculations suggest that the 98% level of difficulty suggested by a number of writers is too low and for UR should be closer to 99% to 99.5%. Hu and Nation (2000)

found that 98% coverage of vocabulary led to 72% comprehension, so possibly for weaker readers, overall comprehension should be higher, but further research is needed.

The study showed that the students were capable of acquiring more than one in ten unknown words; however, the number of unknown words in each book is far less than most researchers suggest as appropriate. For an average acquisition after 14 days for this study of 23% from total available words (48), the students could acquire 11 words from a book like *The Promise*, requiring them to read 90 books to acquire 1000 words or about 2.5 years (ten days to read a book). This is considerably more than the 28 – 56 books over 1.0 – 1.5 years suggested by other researchers (Nation, 2001; Stahl, 1999; Schmitt, 2000; Stahl and Nagy, 2006).

This is a fairly static calculation in that it does not take into account an improved ability to acquire words or a larger total number of available words, making the time frame of four years to acquire 1,000 words quite daunting. However, for the WR employed in the study, using traditional language teaching methods, 50% of them had still not acquired the most common 1,000 words after about twelve years at school studying English. So, while these particular figures show UR to be a slow way to acquire vocabulary,

providing the students were able to actually read the required number of appropriate books, it may still be superior to traditional methods.

While the books represented uninstructed reading, the reading was not unguided in that the students were given the books to read. A problem with giving students a novel to read is that if students do not like the book or it is not at a level that they find appropriate, they still have to read it, which could make reading a burden. An ER programme means that students can self-select topics or genres of interest at an appropriate level, which may raise interest and possibly improve acquisition. However, this again raises the problem of H.K. students' lack of motivation, lack of time and uncertainty as to what to read etc., as noted in Table 4.10. Free choice of materials may not necessarily mean that students will read.

The H.K. extensive reading scheme was based on books and materials recommended by the Edinburgh Project on Extensive Reading (EPER, 1998 – 1.4). Nishizawa et al (2010:633) discuss research which suggests that the lowest EPER reading levels (starter, beginner) were too difficult for low-level Japanese learners to read without translating the English text. This may also be applicable to H.K. Chinese students and might in part explain some of the difficulties encountered with implementing the H.K. extensive

reading scheme (1.6). It is vital that UR is at a level where students may obtain comprehensible input, which this research has indicated may be 99% - 99.5%.

It is, however, unsatisfactory to simply provide students with a box of readers at the 'appropriate level', as within most classes, there may be a range of reading abilities. Students also need to be instructed on how to select a book from those provided, perhaps using methods like the five-finger rule noted above. In a survey of H.K. primary students, 37% of primary 6 students (year 6) said that they did not know how to choose a book (HKIE, 1997:60).

5.2 Research question 2

Can a group of H.K. students acquire vocabulary from reading with tasks?

On the 14-day test, students on average acquired 64.9% from the total words available or 11.23 words, failing to acquire just six available words. The reading tasks exposed the students to the target vocabulary regardless of whether they had read the book or not. It could be argued that the tasks alone, without reading, would have been sufficient for vocabulary acquisition. Huckin and Coady (1999:182), however, suggest that words in

context give the learner a “richer sense of a word’s use and meaning than can be provided in traditional paired-associate type of exercises”.

While this study sought to measure vocabulary acquisition, it is acknowledged that there may be other gains possible from reading, as previously noted (1.4), even where the tasks employed only focused students on vocabulary. One of the vocabulary tasks (R+) was the literature circle, where students discussed words they had found in the story. This activity provided for interaction, which was enriched due to the use of a real book to provide context for the target words. The output demanded by the literature circle meant that a number of times, students asked the teacher for pronunciation of a word. Shared book programmes which incorporate reading with tasks involving an oral component have been shown to improve students’ oral language (Elley and Mangubhai, 1983:61) and listening (Ng and Sullivan, 1991). The interaction was limited due to the nature of the roles given to students which focused primarily on vocabulary. Essentially, a student tried to get their group members to guess a word or word meaning by presenting pictures or acting out a word, etc (Appendix C1). This limited the interaction to a question-answer type of role. Through the use of a variety of roles, literature circles can be highly interactive. An advantage for oral tasks after reading is that they can be difficult to carry out if the reader has not read the book.

Atay and Kurt (2006) noted gains for young Turkish students engaged in reading with interactive tasks over reading with discrete vocabulary study (2.2.5). The interactive tasks were collaborative and involved jigsaw tasks and retelling and rewriting. In the current study, the literature circle was not particularly interactive, as its purpose was primarily to have the students focus on vocabulary as part of the 'study'. Students employed a number of techniques to encourage their group members to guess their target words or to highlight the meanings of their target words, which involved group members in a semi-interactive way. Literature circles, however, can be highly interactive, with students questioning each other about book characters, cultural aspects of a story or a story's deeper meaning. Brown and Campione (2002) suggest that socially interactive learning is critical to learning and propose that learners can play the role of teacher and instigate learning through various book discussions after reading.

There was no formal teaching for the R+ books in the study. Students found word meanings for themselves from the books in order to complete tasks. Answers were provided for the work sheets but there was deliberately no discussion of the answers. The vocabulary learning was explicit, as the learners' attention was directed to a task (Doughty and Williams, 1998), but incidental in that there was no deliberate attempt required for the students to commit new information to memory. The primary purpose

of reading was to enjoy the book by understanding the story, although comprehension was not measured. Hong Kong students are very task orientated. If they are presented with a book to read and a vocabulary work sheet, there is a chance that they will complete the work sheet without reading the book, particularly if the worksheet is to be submitted for marks or grades. Interactive tasks require students to genuinely read the book in order to fully take part. If necessary, part of the interactive process could be assessed if this would further encourage grade-orientated students to read.

The post-reading tasks employed in the current study were all discrete vocabulary-related exercises, which were probably not particularly interesting or motivating. Their primary purpose was to provide students with an opportunity to notice and use the target words. In a genuine classroom reading programme, the tasks could help to develop a variety of skills such as comprehension, writing, oral skill, etc. and could be more motivating and interactive, similar to shared book tasks (Appendix B2) or those used by Atay and Kurt (2006). The tasks could also provide a greater necessity for the book to be actively interacted with and understood in addition to any targeted language aims.

After 28 days, WR had gained 44.7% and SR 44.3% of available words - seven and nine words respectively, or a loss of four words from the 14-day test. As with UR, these results only report on acquisition from the 26 target words and there may have been additional vocabulary and skills acquired. The study does not record partial word knowledge. If words can be retained for 28 days, it increases the chance of their being seen again and their knowledge strengthened.

R+ was presented to students as a packaged programme. The programme aimed to aid vocabulary knowledge by encouraging noticing of unknown words (2.2.2) and an opportunity for output (2.3.0). Tasks aimed to strengthen word knowledge and improve word recognition and background knowledge (schema) were enhanced through the use of pre-reading introductions and literature circles, all seen as problems in weaker readers (Koda, 2005 – see 2.2.2). The study did not isolate individual parts of the package, for example schema development, so it is not possible to comment on their effectiveness. R+ as a whole was shown to be effective in aiding students' vocabulary acquisition.

5.3 Research question 3

From which method can the students acquire the most vocabulary?

Both weaker and stronger readers acquired significantly more vocabulary from reading with tasks than uninstructed reading after 14 and 28 days. Smith (2006) demonstrated larger gains in general language competence for pure ER (UR) over ER+ and intensive reading (IR); however, the gains for ER were almost entirely due to gains in the first semester of the study, with ER+ and IR making considerable gains in the second semester. Smith suggests that the type of supplementation may have an effect on language, with some forms of supplementation being more effective than others (pg. 15).

The ER+ group simply wrote reaction reports of the books read. the reaction reports asked readers to indicate the book read, pages read and give their overall evaluation of the book. A task such as this may require little in the way of original productive language and have a low task loading (Laufer and Hulstijn, 2001). This type of task is not particularly interesting or stimulating for students and for some, could even become a burden to complete after reading. Students with little ability and low motivation may need supplementation that helps to make input easier or more interesting and comprehensible. Interesting, comprehensible input may also encourage students to obtain more input. The significant difference Smith (2006) found between the ER and ER+ group may be

because the ER group spent time acquiring language from reading that the ER+ group spent on a supplementary task (reaction reports) that did not provide them with input. Mason (2004) also concluded that a similar type of supplementation (summary writing) was of little use for language acquisition when added to extensive reading. One should not conclude that supplementation is not beneficial but rather that these particular types of supplementation (summary writing and reaction reports) did not add to the power of reading. Lack of language development from the task is even less surprising when one considers that a large percentage of the students said that they had copied the summaries they wrote (Mason, 2004:11).

In the current study, one may conclude that the vocabulary tasks aided vocabulary acquisition. Language-specific tasks targeting specific language skills supported in context by a book may aid those specific targeted language aspects. Summary writing is often used as a means to check whether a student has read a book and may not have a specific 'language' goal.

Reading and tasks (R+) are able to work in partnership. Individually, they are able to aid language acquisition but they may be more effective combined than individually. Reading can provide a stimulating context for language study. The challenge for the

educator is to provide tasks that develop language skills while being book based and motivating and relevant to students' needs (Doughty, 2003).

One simple explanation for the success of R+ over UR is that more time was spent on R+ and therefore one could expect better results. Mason (2004) found that summary writing combined with reading had a positive effect but not when the time to complete the task was taken into account in comparison with the gains made by groups who did reading only. More time was spent on R+ than UR, producing significantly better vocabulary acquisition. Gains from UR14 over both groups from available words were 22% compared with 65% for R+ after 14 days. R+ required approximately an extra 160 minutes of work per book, of which about 90 minutes was class time. Students gained, on average, about 7 to 8 words for the extra time spent or 1 word for about 22 minutes retained after 14 days where they were not attempting to deliberately remember the words for a test. R+ enabled students to gain words at three times the rate of UR. This gain may appear to not be particularly worthwhile if students could read a book in about an hour, meaning that instead of doing tasks, they could have read two or three more unguided books. The best gain for UR14 was 3.71 words, and if students read three books this would give them about 11.1 words, which is about the same as for R+ (11.2 words). However, Min (2008), in a study of third year secondary school intermediate

learners in Taiwan, found a retention advantage for reading with enhancement tasks over repeated reading for vocabulary acquisition. The study compared reading followed by a variety of vocabulary-enhancing exercises (R+) with repeated thematic reading. Both groups were exposed to the target words a number of times, one through reading and tasks and the other through reading a number of related articles. The study also reported a gain for the R+ group over the repeated reading group after three months, indicating better retention of vocabulary from R+. The study indicates that even where time is controlled, R+ may be more effective for acquisition than UR. Krashen acknowledges that skill building and output plus feedback can lead to consciously learned competence, but simply suggests that acquisition from comprehensible input alone is more efficient (Krashen, 2004c). Krashen's evidence in L2 for this efficiency from ER is scant; however (Table 2.1), especially so when compared with the additional benefits that appropriate supplementation may add to reading through R+ (see 2.3.5 and Appendix B3).

While R+ takes some time away from a student reading more books, it increases the frequency of exposure to targeted unknown words through the use of tasks, making R+ more efficient, at least for vocabulary acquisition. While Waring and Takaki (2003) suggest that reading (UR in the form of ER) helps to enhance and further develop

vocabulary that learners have acquired previously, this type of word building can be rather haphazard, as firstly, reading needs to be extensive, secondly, the word needs to be encountered frequently, and thirdly, while frequent exposure 'may' eventually lead to acquisition, this is not guaranteed. Joe (2010:134), in a case study of a single subject, found that vocabulary learning was cumulative, with a shift from no knowledge to perceived word form familiarity which was developed through frequency of encounters more than through contextual richness. If a student requires ten exposures to an unknown word in order to acquire its form-meaning and the word appears twice in a book, the student would need to read five books. However, the words may not occur in a subsequent book or books for some time, during which time the trace memory of the word may be erased, therefore requiring more than ten actual exposures. R+, while not guaranteeing acquisition, is able to control the number of exposures a student may have to any selected word or the frequency of encounters (Joe, 2010). A word which, for example, only appears twice in a book, can be used in tasks so that students are exposed to it as many times as desired.

While R+ may require more time than UR, the time may be more efficient by ensuring frequent exposure to targeted vocabulary, which is not possible for UR. Nineteen students (including the six non-readers) failed to acquire any vocabulary at all from the

two UR books in the study for both groups compared with only two students for R+ in total (Tables 4.2 and 4.5). R+ would include the same general gains developed from extensive reading (see 1.3; world knowledge, comprehension, etc) as well as additional targeted aims (oral, listening, specific grammar, etc).

R+ could be used more effectively than in the current study to control reading speed or the quantity read over time. A realistic set number of pages could be given to students each day for homework with a quick quiz the next day to monitor the homework and to encourage its completion. In the current study, most students completed three or four chapters out of fifteen in the ten-minute initial class reading time, suggesting that in a sustained session, most students could complete reading the book in an hour. Two or three chapters each night could have the students starting a new book every Monday and the increased reading quantity and less delayed exposure to new vocabulary might significantly increase vocabulary acquisition. While unguided reading could be carried out in a similar fashion, 'checking quizzes' are much more difficult with thirty students all reading different texts, which was a problem the researcher initially found in his H.K. experience.

The Woman in Black (WR book C, SR book D) was read by WR as R+ where they were provided with some pre-reading guidance and during reading and post-reading tasks. Measured after 28 days, they acquired 44.7% of available words (6.93 actual words). Stronger readers read the book uninstructed and were measured after 14 days, and acquired 27.32% of available words (3.5 actual words).

An independent-samples t-test compared the vocabulary gains between WR R+28 $M = 3.50$, $SD = 2.28$) and SR UR14 $M = 6.93$, $SD = 4.27$), $t(28) = 2.69$ $p < .05$ (sig: .014). The magnitude of the difference in the means = 3.43, 95% CI: -6.09 to -.761 with a very strong effect size, $r = .45$. This means that the WR gained significantly more vocabulary from reading a book with tasks measured after 28 days than SR gained from reading the *same* book without tasks measured after only 14 days. Not only did WR acquire more vocabulary from the book but, as noted from Table 3.7, they enjoyed the book more (4.7) than SR (5.2) and found it easier (4.2) than the stronger readers (4.9) although an independent-samples t-test showed little difference (difficulty $r = .21$; enjoyment $r = .12$).

In this example, R+ was able to make a difficult book easier and more enjoyable and to enable weaker students to acquire significantly more vocabulary from it than stronger

readers who read it unaided. The researcher was curious as to whether there was a correlation between perceived difficulty and enjoyment. If students believe that a book is easier, will they also enjoy it more? The opposite is probably true in that if a book were very difficult, there would be little comprehension and presumably little enjoyment.

The relationship between perceived enjoyment of all the books read in the study and perceived difficulty was investigated using Pearson's product-moment correlation coefficient. Preliminary analyses were performed to ensure that there were no violations of normality, linearity and homoscedasticity. There was a strong positive correlation between the two variables, $r = .615$, $n = 96$, $p < .0005$ (.01). The positive correlation means that if a book is perceived as being easier, it will also be perceived to be more enjoyable, although a correlation does not mean causality.

It is suggested from the example of the book *The Woman in Black* (books C and D) that tasks may help to make a book easier, and if students find the book easier, they may also enjoy it more. It is suggested that the students may prefer tasks, as they believe that tasks will help their English more than reading without tasks (4.4.1). Unfortunately, the students were not asked directly which method (UR or R+) they enjoyed the most.

The rate of vocabulary decay between both groups (WR and SR) across UR and R+ was fairly similar. In fact, there was slightly greater decay for R+ than for UR, although the difference was not significant. In this study, vocabulary from two different books was tested after 14 days and 28 days. To more accurately measure this, the decay on the same book would need to be calculated. One would surmise that R+ would give rise to learning that would last longer than from UR. The higher rate of decay from R+, however, may be because the more one learns, the more opportunities there are to forget. For example, SR only acquired on average 3.5 words from UR14 and they lost most of them after 28 days (0.92 remaining). This could be described as a 'floor effect' (the opposite of a ceiling effect) in that the scores were unable to fall lower. However, on R+, students acquired 12.91 words after 14 days, which reduced to 8.63 words after 28 days. For R+, there was a greater opportunity for decline. Keating (2008), in a study of seventy-nine beginning learners of Spanish, also found that vocabulary tasks which produced the greatest vocabulary gains also suffered the greatest loss when tested two weeks after the treatment.

5.4 Research questions 4 and 5

Which method (R+ or UR) do the students prefer? Does either method better aid the weaker or stronger readers?

Weaker readers acquired slightly more vocabulary from uninstructed reading than did stronger readers and the rate of decay was slightly less from 14 to 28 days. It may seem natural at this point to explore the reasons why weaker readers were able to acquire as much vocabulary as stronger readers. If, however, the situation were reversed and the stronger readers had significantly outperformed the weaker readers, one would possibly not see the need to explain why this had happened, as it is expected that SR will outperform WR. Horst et al (1998) predicted that subjects who knew more words generally would find it easier to understand the text and learn new words from it than would subjects with smaller vocabulary sizes, reflecting the 'rich get richer' phenomenon. Few studies have actually attempted to measure group differences in language or vocabulary acquisition, however. Where studies have attempted to make a distinction between learner abilities and their ability to acquire language (Tekmen and Daloğlu, 2006; Zahar et al, 2001; Horst et al, 1998) the same book or passage was employed for both the stronger and weaker readers where, perhaps unsurprisingly, both groups performed according to expectation. In spite of the paucity of studies comparing

ability groups, we expect stronger readers to outperform weaker readers. But if the reading materials are at an appropriate level for both groups, as in the current study, then both groups may perform equally well.

There is no evidence to suggest that the weaker readers in the current study are lacking in intelligence in any way. They may be ‘weaker’ due to a lack of opportunities in earlier years to have acquired language. They may have arrived from the mainland (China) more recently with poorer English. There are any number of reasons why the weaker students have less vocabulary and reading comprehension ability than the SR in the study.

While WR attained a similar number of words from UR 14 as SR, they had more available words from a slightly shorter novel, which they found easier and more enjoyable. A key difference between the books and the ability to acquire vocabulary may also be the perceived difficulty and enjoyment level of the text. WR indicated that they found their books easier and more enjoyable than SR, which may have aided vocabulary acquisition. The results from the initial vocabulary placement tests were shared with the students and weaker readers may have been more conscious of their need to improve. They may have been more motivated by the motivational talk than the SR. As noted earlier, (2.1.3) Wang and Guthrie (2004) suggest a strong relationship

between intrinsic motivation, quantity read and overall comprehension, which may also influence the acquisition of vocabulary.

While WR performed as well on UR as the stronger readers, they expressed a preference for reading with tasks. Qualitative data might have been useful to explore the reasons for this, which for now we can only speculate on. Generally, reading in H.K., or in fact any class work, is conducted with an entire class all working on the same book, textbook or material, which is more often than not assessed. The level of work is often set at one appropriate to the corresponding national examination, which, for students in CMI schools and particularly weaker students, is a level that is generally difficult for many of them. The weaker students in this study may have preferred reading with tasks because they found the scaffolding helpful in enabling them to complete the reading, which is normally seen as a difficult task. WR prefer tasks because they are used to reading books that are usually too difficult and the tasks assist them. Their inclination may reflect a lack of confidence in their ability to handle UR. With a book at an appropriate level, they are quite capable of acquiring language (vocabulary) without the addition of tasks, on which they have perhaps become dependent. Independent, unguided, uninstructed reading may appear somewhat daunting if one is a 'weaker reader' compared with guided reading.

Stronger readers showed a slight non-significant advantage for R+14 over weaker readers (Table 4.3). Both groups read the same book (*Zorro*). While both groups perceived the books similarly in terms of difficulty and enjoyment, the book must have been more difficult for the WR than the SR. WR probably had a greater number of unknown words in addition to the 26 targeted items. The greater quantity of vocabulary may have meant that even with tasks, their mental loading was greater than for the SR, who had more available processing for the target words. While R+ was able to be employed with a class of mixed ability students and tasks were able to aid learners, the text is still not at the same level of difficulty for all learners such that SR may outperform WR on acquisition. Ideally, text still needs to be at an appropriate starting level for all readers to make similar gains. However, using the same book but employing different tasks may aid the WR better by more directly meeting their needs. WR were still able to gain more than ten words from the book (39%) or a 52% gain from total words available.

5.5 Limitations

While every effort was made to match books that were being compared across methods (UR and R+) or between groups (SR and WR), there may still have been unaccountable differences, which may have affected vocabulary acquisition. While target word, verb and noun frequencies were fairly well-matched, the contexts in which the target words were used were not. The genre and writing styles of books varied, as did the students' preferences. Some of these differences may have enabled a method or a group to perform a little better than another, although there is little suggestion in the literature that these differences may have led to significant differences in vocabulary acquisition.

In order to compare the vocabulary acquisition of WR and SR from an uninstructed text, the text needs to be at an appropriate level for both groups, which means different texts. One way to reduce the variables between the texts would be to use the same story written in an easier and more difficult version. This would eliminate genre, style and other content differences, strengthening the claim for group differences.

The claim made for vocabulary acquisition after 14 and 28 days is a minimal time period and may have varied between groups and students. Students had ten days to read a book but they may have completed it in two days or required ten. Students finishing

the book after two days would in fact have had their vocabulary acquisition measured 22 or 36 days after reading was completed. If the SR all completed the book promptly and the WR took the maximum amount of time, then this could affect a comparison made between them on vocabulary acquisition. While it may be reasonable to suggest that SR read faster, their books were also longer. Reading a book quickly would suggest that it is fairly easy and yet the WR found their books easier than the SR. There is no evidence to suggest that one group completed a book significantly earlier than another; however, this is something that a replication of the study could monitor.

The comparisons made between the 14 and 28-day tests compared gains for one book with gains from a different book. Obviously, the best way to measure vocabulary decay is to measure the same words over time, although each time the words are tested provides students with an additional opportunity for an aspect of word knowledge to be reinforced.

Students had no prior notice of the first post-test, which tested the first two books read. However, this test may have alerted them to the possibility of a further test. When the first post-test was administered (Table 3.12, step 7), students had already read the third book, so there was no chance for them to acquire any additional vocabulary from it

unless they had kept a record. The third book was not tested in the first test and students may have reasoned that as the third book, which they had already completed, had not been tested, then that was the end of the tests. A test pattern is not readily established from a single test, so there is no reason to suggest that the students knew that they would be tested on the fourth book they read.

The relatively small sample size of the groups means that small variations in a single-subject's performance may translate into a relatively significant result. With both groups completing the same tasks at the same time, these variations are reduced. The small sample size increases the risk of a Type II error (there is no effect when there is), but this is reduced through the use of effect sizes and confidence intervals which are not bounded by sample size. Confidence in the results is strengthened by their similarity with other studies (5.1.0). The only way to have complete confidence in the results is by replication of the study and every effort has been made to meticulously describe the procedures and the student sample.

This study only measured the most basic form-meaning link between words. While the measurement after 14 and 28 days indicates a relative strength to this relationship, this 'strength' can not be extended into a suggestion that the subjects acquired anything more than a simple, approximate meaning of a word. While the literature suggests that word knowledge from reading is cumulative, we can not state that a gain of 3 or 4 words from a single UR, or R+, will translate into a gain of 15 or 20 words from five books. A similar study measuring productive word knowledge could provide an indication on the depth of knowledge a learner has of an acquired word including form, meaning and use.

CHAPTER 6

CONCLUSION

This study has shown that where a text is at an appropriate level of difficulty, both weaker and stronger readers can acquire vocabulary from uninstructed reading. The evidence from these Chinese year 12 students provides some additional support for Elley's (2001:238) claim that 'increased literacy learning through exposure to high interest books can occur regardless of the student's first language, age or culture of the classroom'. The data from this study only indicates a vocabulary gain from reading for a small sample of students. The sample were non-readers with relatively poor vocabulary knowledge after many years of English language classes. While the gains are small, there is a suggestion that if these small gains could be encouraged through an R+ or an ER programme, increased literacy may result.

While Elley notes that learning *can* occur, it is not guaranteed. The mere provision of interesting books is not sufficient for language growth: students need to interact with the books productively. Students who failed to read the first book failed to acquire any vocabulary. The failure of H.K. ER studies to demonstrate vocabulary gains is most likely to reflect the inability of the programmes to sufficiently engage students. Whether the lack of engagement is the fault of the researcher / teacher or a clash between the principles of ER and local educational pedagogy would need to be determined by further research, but we do know that both weaker and stronger Chinese L2 learners can acquire vocabulary from UR when they interact with a text.

The current study and various other H.K. studies indicate that learning vocabulary from uninstructed reading can be haphazard. Some readers are able to acquire vocabulary and some are not, but acquisition is not necessarily related to reading ability but more to the level of engagement by students. In the current study, the presence of varying degrees of engagement is somewhat understandable in that the stories were not selected by the students, but in Lai (1993a, 1993b) students self-selected books as per an ER programme and yet some still failed to adequately engage with text and acquire language.

In addition to ensuring that readers actively engage with text, there is also the need for the text to be at an appropriate level such that a reader can understand it sufficiently to enjoy its message and that the text provides sufficient support to enable accurate guessing of words. While some writers have suggested a level of 98% comprehensibility, this study found evidence that the figure may need to be closer to 99% or 99.5% for weaker readers. It is vital that text be at a level whereby the reader can enjoy the message, which means comprehending it easily. This study has suggested that text difficulty is correlated with enjoyment, and for weaker or reluctant readers to develop a habit for reading, they must experience enjoyment, particularly if they are to continue with independent reading for years. A quarter of students in the study reported that a reason for their not reading was that reading was too difficult. This would indicate that they do not know how to select books at an appropriate level and this may have been an additional cause of failure in H.K. reading studies. While selecting a book at an appropriate level is basic to ER programmes, further H.K. research could investigate whether this is a component that is lacking in H.K. ER programmes.

It is possibly quite heartening for many teachers of English as a second language around the world that the weaker students in this study were able to acquire some vocabulary from uninstructed reading. While there may be a threshold level of vocabulary knowledge below which a learner cannot read well enough to learn new vocabulary through reading, as suggested by Coady (1997), the results for uninstructed reading from this study suggest that this level, if it exists, is somewhere below the 1,000-word level. There was no correlation between a student's prior vocabulary knowledge and their ability to acquire vocabulary, although the student with the lowest vocabulary knowledge did acquire the least number of words from the four books. Further research is required to determine the exact level of the 'beginner's paradox'. The threshold level may relate more to attitude, motivation and the affective filter than to actual reading ability.

If a programme is able to ensure that the students continue to enjoy reading, then modest gains from a single book can be multiplied over time through an increased number of completed books (ER). While the rate of acquisition is quite low for UR, taking students 2 – 3 years to acquire 1,000 words, this is still superior to the current methods that have been employed for vocabulary learning. The difficulty with UR is engaging students sufficiently, such that small gains may be multiplied over time.

Equally encouraging to the gains from UR are those from R+. While a little more time was required for this programme, the gains were considerably greater than for UR and quite robust even after 28 days. R+ was able to make a book easier and difficulty was correlated positively with enjoyment. R+ therefore shows promise in its ability to raise the enjoyment level of texts, which may aid engagement. Weaker students were able to make greater gains from a book from R+ than were SR students with the same book

uninstructed. Further research is required to investigate the quantity and type of tasks best suited to improving engagement, comprehensibility and enjoyment, without the reading becoming intensive.

One advantage for R+ was that all students in the class were able to read the same book (*Zorro*) at the same time. The tasks (pre, during and post reading) were able to provide sufficient help to make the book easier for the weaker readers. Both groups reported almost the same ratings for the book on difficulty and enjoyment. A large class of students all reading the same book may give a teacher more control over students' reading and the ability to genuinely monitor the reading. With specific knowledge of a book, a teacher is able to briefly confer with (conference) each student to ascertain that reading has actually taken place and gauge the student's ability to cope with the text. Once targeted words have been introduced in context, they can be revisited by the teacher at any time and can also be targeted and revisited as they appear again in subsequent reading. Having a whole class read the same book makes assessment and accountability much easier to manage for a class of 40+ students than if the students all read their own self-selected books and complete reaction reports or summaries which do not provide language input and may be completed without even reading the books.

Having an entire class read the same book or dividing them into just two or three manageable groups could overcome some of the pedagogical / philosophical problems encountered by ER (see 1.6). R+ could be timetabled into the curriculum, overcoming the problem of students (and teachers) having no time for reading and associated work. R+ can reduce the excuses for students not reading by carefully selecting interesting books and tasks. Regular exposure to interesting stories, while developing confidence and skills, may also begin to stimulate enjoyment and a reading habit.

R+ would give a purpose to reading 'homework', thereby encouraging parental support for it. A few elements from a book could be incorporated into tests or examinations as necessary. These could include vocabulary or grammar items, writing tasks describing characters or reactions to the book or oral examination tasks: 'What birthday present would you buy for a character in *The Promise* and why?' More capable students could be extended with more demanding tasks while the weaker students are supported and guided in their reading at the same time.

R+ would mean that reading would remain teacher-centered, which is familiar for students, while providing them with a reading experience which could initially be structured and controlled and later move to less structured and less guided as ER was introduced. R+ could help build confidence with reading while building basic vocabulary that would be necessary later for students attempting ER.

The study has suggested that the frequency of new words encountered influenced acquisition for WR. With UR in the form of ER, frequency is uncontrolled and is determined by quantity. Weaker, reluctant or slower readers may not be able to carry out sufficient reading to ensure adequate repetition. It may be possible for a teacher to cajole a student through a reading programme for a year but they need to keep reading for several years to acquire the necessary basic vocabulary. With R+, frequency of repetition can be controlled, thus ensuring sufficient repetition of targeted words, better guaranteeing acquisition and making R+ efficient and effective.

The danger with tasks is that they can become more important than the reading or be completed without the need to read. Often they are employed simply as a means to

monitor reading. The written tasks in this study were designed primarily as a means to reinforce the target vocabulary for the purpose of the experiment. Tasks could be designed that are closely linked to the story content such that the student would genuinely need to read the book in order to complete them, allowing the tasks to both monitor readers and fulfil a linguistic purpose. Tasks could be interactive where necessary, or could even involve students working online or producing PowerPoint presentations and book talks. With careful design, interactive tasks could enhance motivation to read, although further research is needed to determine which tasks may be able to accomplish this best for the age and level of the students. While the current study only focused on vocabulary, any skill could be developed in context in a systematic way as desired.

Students in H.K. with little experience with ER may need an R+ programme to help them to build some basic vocabulary, develop a reading habit and enjoyment of books and to build reading confidence and independence before launching into UR. In Fiji in the 1980s, shared reading ensured that the students interacted constructively with the books every day and gradually built up their knowledge of the language (Elley, 2000:237). After students had developed confidence in reading, and possibly a reading habit, independent reading (UR) was introduced later in the year ‘when the pupils were ready’ (Elley, 2000:240).

While this study has compared and contrasted weaker with stronger readers, the ability of the two groups is relative. Further research is required to ascertain if these results can be generalised to considerably weaker and stronger students. What should be noted from the book flood studies by Elley and colleagues is that their programmes were generally employed in primary schools. With higher intrinsic motivation to read and a lower affective filter, a genuine readiness to read for pleasure may be easier to develop at a younger age.

This study has only measured vocabulary acquisition in its simplest form-meaning relationship. It is hoped that ultimately reading would lead to an improved ability to use the newly acquired vocabulary productively. R+, which presents vocabulary both implicitly and explicitly, may more effectively develop productive language than UR which relies on implicit acquisition alone. While we may suggest from the results of the study that reading is our answer to solving the language problems of H.K. students (1.4), we must bear in mind that the study has only measured a rather tentative form-meaning relationship from a small sample of students. Further research is required to measure productive acquisition from R+ and UR and between stronger and weaker readers.

There is no suggestion that R+ should be a permanent programme in H.K. schools or the only reading programme. Ultimately the aim would be to develop independent readers who can read for enjoyment. While this study demonstrated vocabulary gains in a controlled experimental situation and suggested a role for R+ in developing ER, actually maintaining a regular ER programme, particularly with older students, is not guaranteed. Further research is required to determine the extent to which R+ may be able to develop reading habits and skills in students to ensure that they are 'ready' for independent

reading in the form of ER or the ability of R+ to scaffold UR, and at what stage R+ could be dispensed with, leaving behind a viable ER programme.

This study is by no means concluded. The first part of the study, which posed a problem and decided to test a solution, has been concluded. We now know that H.K. L2 students can acquire vocabulary from UR. We can suggest that problems with ER in H.K. are related to implementing the programme satisfactorily rather than to an inability of the students to acquire language from reading. We know that lack of interest and motivation for reading, a lack of time and the inability to choose suitable books are major factors preventing students from engaging with books. The study has shown that R+ is capable of making reading easier, thus aiding enjoyment, which must go some way towards improving motivation. R+ is also able to provide time and purpose for students to read. We have discovered that R+ is better able to guarantee acquisition of vocabulary from reading than UR, with larger quantities of acquired words, making R+ more effective and efficient for less capable learners.

This project will continue with a sharing of the results with participants and colleagues and a discussion on the way forward. For now, the study has demonstrated a favourable effect for R+ but it remains to be seen whether such a programme can fully overcome the implementation and cultural / philosophical problems encountered by ER and help in developing students as independent readers.

REFERENCES

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), 382 -401.

Ambatchew, M.D. (2004) *The effect of primary English readers on reading skills in Ethiopia: A study in African educational needs*, unpublished paper for D.Litt. dissertation, University of Pretoria, South Africa, available from ProQuest Digital Dissertations database (UMI Publication No. AAT 0807900), <http://upetd.up.ac.za/thesis/available/etd-06142004-143840/unrestricted/00front.pdf> Online, accessed May 28 2007.

Anderson, R.C. and Pearson, P.D. (1984) A schema-theoretic view of basic processes in reading comprehension, in Pearson, P. D. (ed.), *Handbook of Reading Research*, 255 - 291, New York: Longman.

Anon (2009) Brownstone Intermediate School suggested summer reading, (<http://www.portlandlibraryct.org/PDF/Summer2009/SchoolLists2009/bis2009.pdf>) Online, accessed 2 Sept 2010.

Aranha, M. (1985) Sustained silent reading goes East, *Reading Teacher*, 39 (2), 14-217.

Atay, D. and Kurt, G. (2006) Elementary school EFL learners' vocabulary learning: The effects of post-reading activities, *Canadian Modern Language Review / La Revue Canadienne des Langues Vivantes*, 63 (2), December, 255-273.

Balla, J. Stokes, M. and Stafford, K. (1991) Changes in student approaches to study at CPHK: A 3-year longitudinal study. In Conference Proceedings of the *Australasian Association for Institutional Research*, 7-31, Melbourne: AAIR.

Beck, I.L. and McKeown, M.G. (1989) Expository text for young readers. The issue of coherence, in Resnick, L. (ed.) *Essays in honor of Robert Glaser, Insightful overview of text characteristics that reduce coherence*, 47-66, Hillsdale, NJ: Erlbaum.

Beck, I.L., McKeown, M.G. and Worthy, J. (1995) Giving a text voice can improve students' understanding. Empirical study of the effects of giving a text "voice," includes sample materials illustrating the concept, *Reading Research Quarterly*, 30 (2), 220-238.

Beck, I.L., Perfetti, C.A. and McKeown, M.G. (1982) The effects of long-term vocabulary instruction on lexical access and reading comprehension, *Journal of Educational Psychology* 74 (4), 506-521.

Becker, L.A. (2000) *Effect Sizes (ES)*, (<http://web.uccs.edu/lbecker/Psy590/es.htm>) Online, accessed 13 November 2009.

Bell, T. (2001) Extensive reading: Speed and comprehension, *The Reading Matrix*, 1 (1), 1-13.

Bensoussan, M. and Laufer, B. (1984) Lexical guessing in context in EFL reading comprehension, *Journal of Research in Reading*, 7 (1), 15-32.

Bernhardt, E.B. and Kamil, M.L. (1995) Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses, *Applied Linguistics*, 16 (1), 15-34.

Bialystok, E. (1997) Effects of bilingualism and biliteracy on children's emerging concepts of print, *Developmental Psychology*, 33 (3), 47-60.

Birch, B.M. (2007) *English L2 Reading. Getting To The Bottom*, New Jersey: Laurence Erlbaum Associates Inc.

Bishop, H. (2004) Noticing Formulaic Sequences - A Problem of Measuring the Subjective, LSO, *Working Papers in Linguistics* 4: 15-19. Madison, Wisconsin: University of Wisconsin-Madison, Department of Linguistics, Proceedings of WIGL 2004.

Bond, O.F. (1926) Reading for language power, *The Modern Language Journal*, 10 (7), 411-426.

Braine, G. (2009) *Practicing what we preach*, The Chinese University of Hong Kong ([http://nnest-evo2009.pbworks.com/f/Practicing_what_we_preach_\(updated\).doc](http://nnest-evo2009.pbworks.com/f/Practicing_what_we_preach_(updated).doc)) Online, accessed 28 July 2009.

Brown, A.L. and Campione, J.C. (2002) Communities of learning and thinking, in a context by any other name, in Wood, P. (ed.) *Contemporary Issues in Teaching and Learning*, Falmer, Routledge.

Brown, J.D. (1998) An EFL readability index, *JALT Journal*, 20 (2), 7-36.

Brown, R. (2000) Extensive Reading in the EFL Class (Extensive Reading in Practice Colloquium), *JALT International Conference*, Shizuoka,
(<http://www.extensivereading.net/er/rbrown.html>) Online, accessed July 26, 2010.

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

Carrell, P.L. (1983) Some issues in studying the role of schemata or background knowledge in second language comprehension, *Reading in a Foreign Language*, 1 (2), 81-92.

Carrell, P.L. (1985) Facilitating ESL reading by teaching text structure, *TESOL Quarterly*, 19 (4), 727-752.

Carrell, P.L. (1987) Content and formal schemata in ESL reading, *TESOL Quarterly*, 21 (3), 461-481.

Carrell, P.L. (1988) Interactive text processing: Implications for ESL/ second language reading classrooms, in Carrell, P., Devine, J. and Eskey, D. (eds.) *Interactive approaches to second language reading*, 239-259, New York: Cambridge University Press.

Carrell, P.L. and Eisterhold, J.C. (1983) Schema theory and ESL reading pedagogy, *TESOL Quarterly*, 17 (4), 553-569.

Carroll, J.B. (1972) Defining Language Comprehension: Some Speculations, in Freedle, R.O. and Carroll, J.D. (eds.), *Language comprehension and the acquisition of knowledge*, 1 – 29, Washington, D.C.: V.H. Winston and Sons.

Carter, D.F. and Hurtado, S. (2007) Bridging key research dilemmas: Quantitative research using a critical eye, *New Directions for Institutional Research*, 133 (Spring), 25-35.

Carver, R.P. (1994) Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction, *Journal of Reading Behavior*, 26 (4), 413-43.

Cheung, D.C.K. (1990) The Language of Success, *The South China Morning Post*, January 15th.

Chew, L.C. (2003), *Extensive Reading in English: Hong Kong secondary one students' response to high-interest unsimplified stories*, unpublished paper for Master of Education, University of Hong Kong.

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

Cleeremans, A.A., Destrebecqz, A.A. and Boyer. M.B. (1998) Implicit learning: news from the front, *Trends in Cognitive Sciences*, 2 (10), 406-416.

Coady, J. (1979) A psycholinguistic model of the ESL reader, in Mackay, R.B., Barkman, B. and Jordan, R. (eds.), *Reading in a second language: Hypotheses, organization and practice*, 5-12, Rowley, MA: Newbury House

Coady, J. (1997) L2 vocabulary acquisition through extensive reading, in Coady, J. and Huckin, T. (eds.), *Second Language Vocabulary Acquisition*, 225-237, Cambridge: Cambridge University Press.

Coady, J. and Huckin, T. (1997), *Second Language Vocabulary Acquisition*, Cambridge: Cambridge University Press.

Coady, J. and Nation, P. (1988) Vocabulary and reading, in Carter, R. and McCarthy, M. (eds.) *Vocabulary and language teaching*, 97 – 110, London: Longman.

Cobb, T. (2007) Computing the vocabulary demands of l2 reading, *Language Learning and Technology*, 11 (3), 38-63, (<http://lt.msu.edu/vol11num3/cobb/>) Online, accessed 18 August 2010.

Cobb, T. (2008) Web Vocabprofile, an adaptation of Heatley and Nation's (1994) Range, (<http://www.lex tutor.ca/vp/>) Online, accessed 18 August 2009.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences* (2nd edn.), Hillsdale, NJ: Erlbaum.

Cohen, J. (1990) Things I Have Learned (so far), *American Psychologist*, 45 (12), 1304-1312.

Cohen, J. (1992) A power primer, *Psychological Bulletin*, 112 (1), 155-159.

Cohen, L., Manion, L. and Morrison, K. (2007) *Research Methods in Education* (6th ed.), London, NY: Routledge.

Corson, D.J. (1997) The learning and use of academic English words, *Language Learning*, 47 (4), 671-718.

Coxhead, A. (2000) *Academic Word List*, (<http://www.uefap.com/vocab/select/awl.htm>) Online, accessed February 23 2008.

Dalle, T. (1988) Reading journals: Solving the problem of input in the ESL composition class, *TESL Reporter*, 21 (2), 23-26.

Darr, C., McDowall, S., Ferral, H., Twist, J. and Watson, V. (2008) *Progressive Achievement Test: Reading*, Wellington, NZCER Press.

David, M.K. (2009) Moving beyond schema: Selecting texts for EFL readers, in Cirocki, A. (ed.), *Extensive Reading in English Language Teaching*, 171-178, Munich, Germany: Lincom.

Davis, C. (1995) Extensive reading: an expensive extravagance, *ELT Journal*, 49 (4), 329 - 36.

Day, R. (2009) *The Benefits of Extensive Reading (ER)*, (<http://www.oupe.es/es/ELT/Readers/bookworms/Recursos%20Destacados/the%20benefits%20of%20extensive%20reading.pdf>) Online, accessed 4 March 2010.

Day, R. and Bamford, J. (1998) *Extensive Reading in the Second Language Classroom*, Cambridge: Cambridge University Press.

Day, R. and Bamford, J. (2000) Reaching reluctant readers, *English Teaching Forum*, 38 (3), (<http://exchanges.state.gov/forum/vols/vol38/no3/p12.htm>) Online, accessed 5 January 2008.

Day, R. and Bamford, J. (2002) Top ten principles for teaching extensive reading, *Reading in a Foreign Language*, 14 (2), 136-141.

Day, R. and Bamford, J. (2004) *Extensive Reading Activities for Teaching Language*, Cambridge: Cambridge University Press.

Day, R.R., Omura, C. and Hiramatsu, M. (1991) Incidental EFL vocabulary learning and reading, *Reading in a Foreign Language*, 7 (2), 541-551.

De'Ath, P. (2001) The Niue literacy experiment, *International Journal of Educational Research*, 35 (2), 137-146.

DeKeyser, R.M. (1997) Beyond explicit rule learning: Automatizing second language morphosyntax, *Studies in Second Language Acquisition*, 19 (2), 195-221.

DeKeyser, R.M. (2000) The robustness of critical period effects in second language acquisition, *Studies in Second Language Acquisition*, 22 (4), 499-533.

DeKeyser, R.M. (2003) Implicit and explicit learning, in Doughty, C.J. and Long, M.H. (eds.) *The handbook of second language acquisition*, 313-348, Malden, MA: Blackwell.

Dörnyei, Z. (2005) *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, N.J.: Lawrence Erlbaum.

Doughty, C. (2003) Instructed SLA: Constraints, compensation, and enhancement, in Long, M. and Doughty, C. (eds.) *Handbook of second language acquisition*, 257-310, Malden, MA: Blackwell.

Doughty, C. and Williams, J. (1998) Pedagogical choices in focus on form, in Doughty, C. and Williams, J. (eds.) *Focus on form in classroom second language acquisition*, 197-261. Mahwah, NJ: Lawrence Erlbaum Associates.

Dupuy, B. (1997) Voices from the classroom: Students favor extensive reading over grammar instruction and practice, and give their reasons, *Applied Language Learning*, 8 (2), 253-261.

Dupuy, B. and Krashen, S.D. (1993) Incidental vocabulary acquisition in French as a foreign language, *Applied Language Learning*, 4 (1 and 2), 55-63.

Dykes, G.A. (2007) *Extensive reading in Hong Kong secondary schools*, unpublished paper for the Doctor of Education programme, University of Leicester, School of Education, 28 August.

Dykes, G.A. (2008a) *Reading and implicit and explicit language acquisition*, unpublished paper for the Doctor of Education programme, University of Leicester, School of Education, 7 January.

Dykes, G.A. (2008b) A brief review of data collection and analysis methods and a piloted quasi-experiment, unpublished paper for the Doctor of Education programme, University of Leicester, School of Education, 7 January.

Education Commission (1990) *Education Commission Report*, 4 (ECR 4), Hong Kong: Government Printer.

Elley, W.B. (1991) Acquiring literacy in a second language: The effect of book-based programs, *Language Learning*, 4 (3), 375-411.

Elley, W.B. (2000) The potential of book floods for raising literacy levels, *International Review of Education – Internationale Zeitschrift für Erziehungswissenschaft – Revue Internationale de l'Education* 46 (3/4), 233-255, Netherlands: Kluwer Academic Publishers.

Elley, W.B. (2001) Conclusion: what have we learned?, *International Journal of Educational Research*, 35 (2), 237-246.

Elley, W.B. and Croft, A.C. (1989) *Assessing the difficulty of reading materials: The noun frequency method*, Noun Count Software, True Blue Software: NZCER.

Elley, W.B. and Mangubhai, F. (1981) *The Impact of a Book Flood in Fiji Primary Schools*. Wellington: New Zealand Council for Educational Research.

Elley, W.B. and Mangubhai, F. (1983) The impact of reading on second language learning, *Reading Research Quarterly*, 19 (1), 53-67.

Ellis, R. (2001) Second Language Acquisition: The Role of Consciousness, paper presented at the *National Chengchi University*, Taiwan, September 2001, (<http://english.nccu.edu.tw/academic/RodEllis/Lect1.doc>) Online, accessed 13 October 2007.

Ellis, N. and Beaton, A. (1993) Psycholinguistic determinants of foreign language vocabulary learning, *Language Learning*, 43 (4), 559-617.

EPER (1998) *List of graded readers, Current titles*, IALS: University of Edinburgh (<http://www.ials.ed.ac.uk/epermenu.html>) Online, accessed 12 December 2007.

Eskey, D. (2005) Reading in a second language, in Hinkel, E. (ed.) *Handbook of research in second language teaching and learning*, 563-580, Mahwah, NJ: Erlbaum.

Field, A. (2009) *Discovering Statistics Using SPSS* (3rd edn.), London, Sage Publications Ltd.

Floyd, P., and Carrell, P. (1987) Effects on ESL reading of teaching cultural content schemata, *Language Learning*, 37 (1), 89-108.

Folse, K.S. (2004) *Vocabulary myths: Applying Second Language Research to Classroom Teaching*, Ann Arbor, MI: The University of Michigan Press.

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

Fraser, C.A. (1999) Lexical processing strategy use and vocabulary learning through reading, *Studies in Second Language Acquisition*, 21 (2), 225-241.

Friedman, H. (1968) Magnitude of experimental effect and a table for its rapid estimation, *Psychological Bulletin*, 70, 245-251.

Fry, E. (1981) Graphical literacy, *Journal of Reading*, 24, 383-389.

Gass, S.M. and Selinker, L. (2001) *Second Language Acquisition: An Introductory Course*: New York, Laurence Erlbaum Associates.

Goodman, K.S. (1968) The psycholinguistic nature of the reading process, in Goodman, K.S. (ed.) *The psycholinguistic nature of the reading process*, 13-26, Detroit, MI: Wayne State University Press.

Grabe, W. (2000) Reading research and its implications for reading assessment, in Kunnan, A. (ed.) *Fairness and Validation in Language Assessment*, 226-260, Cambridge: Cambridge University Press.

Grabe, W. (2009) *Reading in a Second Language, Moving From Theory to Practice*, Cambridge: Cambridge University Press.

Grabe, W. and Stoller, F. (2002) *Teaching and Researching Reading*, Harlow, UK: Longman.

Gray, W.S. (1967) *Gray Oral Reading Tests*, Indianapolis, IN: Bobbs-Merrill.

Green, C. (2005) Integrating extensive reading in the task-based curriculum, *ELT Journal*, 59 (4), 306-311.

Greenfield, J. (2003) The Miyazaki EFL readability index, *Comparative Culture*, 9, 41-49.

Gregg, K. (1984) Krashen's Monitor and Occam's Razor, *Applied Linguistics*, 5 (2), 79-100.

Griffin, G. and Harley, T.A. (1996) List Learning of Second Language Vocabulary, *Applied Psycholinguistics*; 17 (4), 443-460.

Grobe, J.A. (1970) Reading rate and study time demands on secondary students, *Journal of Reading*, 13, 286-288; 316.

Gu, Y. (2003) Vocabulary learning in a second language: Person, task, context and strategies, *TESL-EJ*, 7 (2), (<http://www-writing.berkeley.edu/TESL-EJ/ej26/a4.html>) Online, accessed 5 June 2008.

Gu, Y. and Johnson, R. (1996) Vocabulary learning strategies and language learning outcomes, *Language Learning*, 46 (4), 643-679.

Guthrie, J. and Wigfield, A. (2000) Engagement and motivation in reading, in Kamil, M.L., Mosenthal, P.B., Pearson, P.D. and Barr, R. (eds.) *Handbook of reading research*, 3, 403 – 22, Mahwah, NJ: L. Erlbaum.

Haastrup, K. (1991) *Lexical Inferencing Procedures or Talking About Words*, Tübingen: Gunter Narr.

Hafiz, F.M. and Tudor, I. (1989) Extensive reading and the development of language skills, *ELT Journal*, 34 (1), 5-13.

Hafiz, F.M. and Tudor, I. (1990) Graded readers as an input medium in L2 learning, *System*, 18 (1), 31-42.

Hagboldt, P. (1925) Experimenting with first year college German, *The Modern Language Journal*, 9 (5), 293-305.

Hall, S.J. (1992) Using split information tasks to learn mathematics vocabulary, *Guidelines*, 14 (1), 72-77.

Hamp-Lyons, S. (1983) Developing a course to teach extensive reading skills to university bound ESL learners, *System*, 11 (3), 303-312.

Haynes, M. (1993) Patterns and perils of guessing in second language reading, in Huckin, T., Haynes, M. and Coady, J. (eds.) *Second language reading and vocabulary*, 46-64, Norwood, NJ: Ablex.

Heatley, A. and Nation, P. (1994) *Range*, Victoria University of Wellington, NZ, (Computer program, available at <http://www.vuw.ac.nz/lals/>).

Heinsman, D.T. and Shadish, W.R. (1996) Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods*, 1 (2), 154-169.

Henrikse, B., Mees, I.M. and Poulsen, E. (eds.) *Perspectives on Foreign and Second Language Pedagogy*, 191–200. Odense, Denmark: Odense University Press.

Hill, S. (2005) *The Woman in Black*, Oxford: MacMillan Heinemann ELT.

Holdaway, D. (1979) *The Foundations of Literacy*, Auckland: Ashton Scholastic.

Holm, A. and Dodd, B. (1996) The effect of first written language on the acquisition of English literacy, *Cognition*, 59 (2):119-47.

- Hong Kong Institute of Education (1997) *Feasibility Study On Extending Extensive Reading in English To All Levels Of Primary And Secondary Schooling: Final Report*, (<http://libdr1.ied.edu.hk/dspace/handle/2260/1253>) Online, accessed 16 September 2007.
- Horst, M. (2005) Learning L2 vocabulary through extensive reading: A measurement study, *The Canadian Modern Language Review*, 61 (3), 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), 207-223.
- Hu, M. and Nation, P. (2000) Unknown vocabulary density and reading comprehension, *Reading in a Foreign Language*, 13 (1), 403-430.
- Huang, C.C. (1994) University students [sic] vocabulary knowledge, content knowledge and reading comprehension, *Journal of National Tainan Teachers' College*, 38 (1), 125 – 153.
- Huang, C.C. (1999) *The effects of vocabulary knowledge and prior knowledge on reading comprehension of EFL students in Taiwan*, unpublished doctoral dissertation, University of Ohio, Athens, OH.
- Huckin, T. and Coady, J. (1999) Incidental vocabulary acquisition in a second language: A review, *Studies in Second Language Acquisition*, 21 (May), 181-193.
- Hui, P. (2003) Hong Kong children ranked poorly for reading skills, *South China Morning Post*, Editorial, April 10th 2003.
- Hulstijn, J.H. (1992) Retention of inferred and given word meanings: Experiments in incidental vocabulary learning, in Arnaud, P.J. and Bejoint, H. (eds.) *Vocabulary and applied linguistics*, 113-125, London: Macmillan.
- 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, *Modern Language Journal*, 80, 327-339.
- Hulstijn, J. and Laufer, B. (2002) Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition, *Language Learning*, 51 (3), 539-558.

Hulstijn, J.H. and Trompetter, P. (1998) Incidental learning of second language vocabulary in computer-assisted reading and writing tasks, in Albrechtsen, D.B., Henriksen, B., Mees, I.M. and Poulsen, E. (eds.), *Perspectives on foreign and second language pedagogy: Essays presented to Kirsten Haastrup on the occasion of her sixtieth birthday*, 191–200. Odense, Denmark: Odense University Press.

Hunt, A. and Beglar, D. (2005) A framework for developing EFL reading vocabulary, *Reading in a Foreign Language*, 17 (1), 23-59.

Ioup, G. (1984) Two readers comment on Stephen Krashen's Input Hypothesis: Testing the relationship of formal instruction to the Input Hypothesis, *TESOL Quarterly*, 18 (2), 345-350.

Irvine, A. (2006) *Extensive reading and L2 development: a study of Hong Kong secondary learners of English*, unpublished paper for Doctor of Philosophy, University of Edinburgh.

Iwahori, Y. (2008) Developing reading fluency: A study of extensive reading in EFL, *Reading in a Foreign Language*, 20 (1),
(<http://nflrc.hawaii.edu/rfl/April2008/iwahori/iwahori.html>) Online, accessed 16 April 2008.

Izumi, S. (2002) Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization, *Studies in Second Language Acquisition*, 24, 541-577, Cambridge University Press.

Jarrell, D.S. (2003) Extensive reading for weak readers, *Journal of Nagoya Women's University*, 49 (3), 199-205.

Jensen, L. (1986) Advanced reading skills in a comprehensive course, in Dubin, F., Eskey, D. and Grabe, W. (eds.) *Teaching second language reading for academic purposes*, 103 – 124. Reading, MA: Addison-Wesley Publishing.

Joe, A. (2010) The quality and frequency of encounters with vocabulary in an English for Academic Purposes programme, *Reading in a Foreign Language*, 22 (1), 117-136.

Johnson, R.K. (1994) Critical review of literature on language education in Hong Kong, in Education Commission (1995), *Education Commission Report, No 6 – Part 2 (annexes)*. Hong Kong: Government Printer.

Johnson, R.K. and Cheung, Y.S. (1995) *Reading Literacy in Hong Kong*, An IEA World Literacy Project on the Reading proficiency of Hong Kong Students in Chinese and English, Hong Kong: The Hong Kong Polytechnic University.

Johnson, M., Kress, R. and Pikulski, J. (1987) *Informal Reading Inventories* (2nd edn.) Newark, Delaware: International Reading Association.

Kameenui, E.J., Carnine, D.W. and Freschi, R. (1982) Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall, *Reading Research Quarterly*, 17 (3), 367-388.

Keating, G.D. (2008) Task effectiveness and word learning in a second language: The involvement load hypothesis on trial, *Language Teaching Research*, 12 (3), 365 - 386.

Kim, N. (2006a) Reading as Input for Grammar Development in Writing: Is Comprehensible Input Enough? *Issues in EFL*, Spring, 4 (1), 3-22.

Kim, Y. (2006b) Effects of Input Elaboration on vocabulary acquisition through reading by Korean learners of English as a foreign language, *TESOL Quarterly*, 40 (2), 341 - 373.

Knight, S.M. (1994) Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities, *Modern Language Journal*, 78 (3), 285-299.

Koda, K. (1995) Cognitive consequences of L1 and L2 orthographies, in Taylor, I. and Olsen, D. (eds.) *Scripts and literacy*, 311-326, Dordrecht, The Netherlands: Kluwer

Koda, K. (1997) Orthographic knowledge in L2 lexical processing: A cross-linguistic perspective, in Coady, J. and Huckin, T. (eds.) *Second language vocabulary acquisition*, 35-52, Cambridge: Cambridge University Press.

Koda, K. (1998) The role of phonemic awareness in L2 reading, *Second Language Research*, 14 (2), 194-215.

Koda, K. (2005) *Insights into Second Language Reading: A Cross-Linguistic Approach*, Cambridge University Press.

Koda, K. (2007) Reading and Language Learning: Crosslinguistic constraints on second language reading development in Koda, K. (ed.), *Reading and Language Learning*, Special issue of Language Learning Supplement 57, 1-44.

Krashen, S.D. (1976) Formal and informal linguistic environments in language acquisition and language learning, *TESOL Quarterly*, 10 (2) 157-168.

Krashen, S.D. (1981) A case for narrow reading, *TESOL Newsletter*, 15 (6), 23-31.

Krashen, S.D. (1982) *Principles and Practice in Second-Language Acquisition*, Oxford: Pergamon Press.

Krashen, S.D. (1985) *The Input Hypothesis: Issues and Implications*. London: Longman.

Krashen, S.D. (1989) We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis, *The Modern Language Journal*, 73 (4), 440-464.

Krashen, S.D. (1993) *The Power of Reading: Insights From The Research*. Englewood, CO: Libraries Unlimited, Inc.

Krashen, S.D. (1998) Comprehensible output? *System*, 26 (2), 175-182.

Krashen, S.D. (2003) *Explorations in language acquisition and use: The Taipei lectures*. Portsmouth: Heinemann,
(<http://books.heinemann.com/shared/onlineresources/E00554/chapter2.pdf>) Online,
accessed 23 July, 2007.

Krashen, S.D. (2004a) Applying the comprehension hypothesis: Some suggestions, paper presented at the 13th *International Symposium and Book Fair on Language Teaching (English Teachers Association of the Republic of China)*, Taipei, Taiwan. 13 November, 2004.

Krashen S.D. (2004b) *The Power of Reading: Insights from the Research* (2nd edn.) USA: Heinemann.

Krashen S.D. (2004c) Why support a delayed-gratification approach to language education? *The Language Teacher*, 28 (7), 3-7.

Krashen, S.D. and Terrell, T.D. (1983) *The Natural Approach*, Oxford: Pergamon Press.

Kuruppu, L. (2001) The “books in schools” project in Sri Lanka, *International Journal of Educational Research*, 35 (2), 181-191.

Kwan, K.M. (1988) Project report on extensive reading, *Hong Kong Reading Association* (quoted in Chew, L.C., 2003), *Extensive reading in English: Hong Kong secondary one students’ response to high-interest unsimplified stories*, unpublished paper for Master of Education, University of Hong Kong,

Kweon, S. and Kim, H. (2008) Beyond raw frequency: Incidental vocabulary acquisition in extensive reading, *Reading in a Foreign Language*, 20 (2), 191-215.

Lai, F.K. (1991) *Extensive reading as input for second language acquisition*, unpublished paper for Doctor of Philosophy, University of Hong Kong.

Lai, F.K. (1993a) Effect of extensive reading on English learning in Hong Kong, *CUHK Educational Journal*, 21 (1), 23-36.

Lai, F.K. (1993b) The effects of a summer reading course on reading and writing skills. *System*, 21 (1), 87-100.

Laufer, B. (1986) Possible changes in attitude towards vocabulary acquisition research, *International Review of Applied Linguistics*, 24 (1), 69-75.

Laufer, B. (1989) What percentage of text-lexis is essential for comprehension?, in Lauren, C. and Nordman, M. (eds.) *Special Language: From Humans Thinking to Thinking Machines*, Clevedon: Multilingual Matters.

Laufer, B. (1992) How much lexis is necessary for reading comprehension? In Arnaud, P.J.L. and Bejoint, H. (eds.) *Vocabulary and Applied Linguistics*, 126-132, London: Macmillan.

Laufer, B. (1997) What's in a word that makes it hard or easy?: Some intralexical factors that affect the learning of words, in Schmitt. N. and McCarthy, M. (eds.) *Vocabulary: Description, Acquisition and Pedagogy*, 140-155, Cambridge: Cambridge University Press.

Laufer, B. (2000) Task effect on instructed vocabulary learning: The hypothesis of ‘involvement’. *Selected Papers from AILA '99 Tokyo*, 47–62, Tokyo: Waseda University Press.

Laufer, B. (2003) Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence, *Canadian Modern Language Review*, 59 (4), 567-587.

Laufer, B. (2005) Ten best ideas for teaching vocabulary. *The Language Teacher*, 29 (7), 3-4.

Laufer, B. (2006) Comparing focus on Form and Focus on FormS in second language vocabulary learning. *Canadian Modern Language Review*, 63 (1), 149-166.

Laufer, B. and Hulstijn, B. (2001) Incidental vocabulary acquisition in a second language: The constraint of task induced involvement, *Applied Linguistics*, 22 (1), 1-26.

Laufer, B. and Nation, P. (1999) A vocabulary-size test of controlled productive ability, *Language Testing*, 16 (1), 33-51.

Laufer, B. and Ravenhorst-Kalovski, G.C. (2010) Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension, *Reading in a Foreign Language*, 22 (1), 15-30.

Laufer, B. and Sim, D. (1985) An attempt to measure the threshold of competence for reading comprehension, *Foreign Language Annals*, 18 (5), 405-411.

Laufer, B. and Yano, Y. (2001) Understanding unfamiliar words in a text: Do L2 learners understand how much they don't understand?, *Reading in a Foreign Language*, 13 (2), 549-566.

Lee, S.K. (2007a) Revelations from three consecutive studies on extensive reading. *RELC Journal*, 38 (2) 150-170.

Lee, S.K. (2007b) Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form, *Language Learning*, 57 (1), 87-118.

Legco Panel on Education (1998) Teaching of English Language in Schools; Progress of Measures to Enhance the Teaching of English Language in Schools, *Paper No. CB(2)621/98-99(04)*,

(<http://www.legco.gov.hk/yr98-99/english/panels/ed/papers/p621e04.pdf>)

Online, accessed 28 November 2008.

- Leung, C.Y.F. (2005) *Developing the culture of reading: A case study of the implementation of reading activities in an EMI secondary school in Hong Kong*, unpublished paper for Master of Arts in Applied Linguistics at the University of Hong Kong.
- Lewis, M. (2000) Learning in a lexical approach, in Lewis, M. (ed.) *Teaching collocation: Further development in the lexical approach*, 155 – 185, Hove: Language Teaching Publications.
- Lin, A.M.Y. (2001) Resistance and creativity in English reading lessons in Hong Kong, in Comber, B., and Simpson, A. (eds.) *Negotiating Critical Literacies in Classrooms*, 83-99, Mahwah, New Jersey: Lawrence Erlbaum.
- Lituanas, P.M., Jacobs, G.M. and Renandya, W.A. (2001) An investigation of extensive reading with remedial students in a Philippines secondary school, *International Journal of Educational Research*, 35 (2), 217-225.
- Liu, N. and Nation, I.S.P. (1985) Factors affecting guessing vocabulary in context, *RELJ Journal* 16 (1), 33-42.
- London, J. (2008) *White Fang*, MacMillan, Oxford: Heinemann ELT.
- Long, M.H. (1983) Does second language instruction make a difference? A review of the research, *TESOL Quarterly*, 17 (3), 359-382.
- Lupescu, S. and Day, R.R. (1993) Reading, dictionaries, and vocabulary learning, *Language Learning*, 43 (2), 263-287.
- Maley, A. (2005) Review of extensive reading activities for the second language classroom, *ELT Journal*, 59 (4), 354-5.
- Mangubhai, F. (2001) Book Floods and Comprehensible Input Floods: Providing Ideal Conditions for Second Language Acquisition, *International Journal of Educational Research*, 35 (2) 147-156.
- Manning, G. and Manning, M. (1984) What models of recreational reading make a difference?, *Reading World*, 23 (May), 375-380.

Mason, B. (2004) The effect of adding supplementary writing to an extensive reading program, *International Journal of Foreign Language Teaching*, 1 (1), 2-16.

Mason, B. (2006) Free voluntary reading and autonomy in second language acquisition: Improving TOEFL scores from reading alone, *The International Journal of Foreign Language Teaching*, 2 (1), 2-5, (<http://www.tprstories.com/ijflt/IJFLTWinter06.pdf>) Online, accessed 3 June 2008.

Mason, B. and Krashen, S.D. (1997) Extensive reading in English as a foreign language. *System*, 25 (1), 91-102.

Mason, B. and Krashen, S.D. (2004) Is form-focused vocabulary instruction worthwhile?, *RELC*, 35 (2), 179-185.

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, England: Multilingual Matters Ltd.

Matsumura, M. (1987) *Eigo no Riidingu (English Reading)*, Tokyo: Taishukan Shoten (2nd edn.) in Susser, B. and Robb, T. (1990) EFL extensive reading instruction: Research and procedure, *The Language Teacher*, (<http://www.kyoto-su.ac.jp/~trobb/sussrobb.html>) Online, accessed 24 March 2008.

McCulley, J. (2007) *The Mark of Zorro*, Oxford: MacMillan Heinemann ELT.

McKenna, M. (2001) Development of reading attitudes, in Verhoeven, L. and Snow, C. (eds.) *Literacy and motivation: Reading in individuals and groups*, 135 – 58, Mahwah, NJ: Lawrence Erlbaum.

McKeown, M.G., Beck, I.L., Sinatra, G.M. and Loxterman, J.A. (1992) The contribution of prior knowledge and coherent text to comprehension, *Reading Research Quarterly*, 27 (1), 78-93.

McLaughlin, B. (1987) *Theories of Second Language Acquisition*, London: Edward Arnold.

McQuillan, J. (1994) Reading versus grammar: What students think is pleasurable for language acquisition, *Applied Language Learning*, 5 (2), 95-100.

McQuillan, J. and Krashen, S.D. (2008) Commentary: Can free reading take you all the way? A response to Cobb (2007), *Language Learning and Technology*, 12 (1), 104-108.

Meara, P. (1996) The dimensions of lexical competence, in Brown, G., Malmkjaer, K. and Williams, J. (eds.) *Performance and competence in second language acquisition*, 35–53, Cambridge: Cambridge University Press.

Meara, P. and Buxton, B. (1987) An alternative to multiple choice vocabulary tests, *Language Testing*, 4 (2), 142-154.

Min, H.T. (2008) EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading, *Language Learning*, 58 (1), 73-115.

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), 249-267.

Monroe, W.S. (1900) *Comenius And The Beginning Of Education Reform*, New York: Charles Scribner's Sons,
(http://openlibrary.org/works/OL996185W/Comenius_and_the_beginnings_of_educational_reform) Online, accessed 8 December 2010.

Morgan, J. and Rinvolucris, M. (1986) *Vocabulary*, Oxford: Oxford University Press.

Morris, P. (1985) Teachers' perceptions of the barriers to the implementation of a pedagogic innovation: A South East Asia case study. *International Review of Education – Internationale Zeitschrift für Erziehungswissenschaft – Revue Internationale de l'Education* XXXI, 3-18. Unesco Institute for Education, Hamburg and Martinus: Nijhoff Publishers, Dordrecht.

Mosenthal, J., Lipson, M., Mekkelsen, J., Russ, B. and Sortino, S. (2001) *Elementary Schools Where Students Succeed in Reading*, The Northeast and Islands Regional Educational Laboratory, a program of The Education Alliance at Brown University, Brown University, (<http://www.alliance.brown.edu/pubs/uvm/elemschread.pdf>) Online, accessed 2 September 2010.

Nadarajan, S. (2009) The effect of instruction and context on L2 learners' vocabulary development, *Electronic Journal of Foreign Language Teaching*, 6 (2), 177-189, (<http://e-flt.nus.edu.sg/v6n22009/nadarajan.htm>) Online, accessed 16 September 2010.

Nagy, W.E. (1988) *Teaching vocabulary to improve reading comprehension*, Newark, DEL: International Reading Association.

Nagy, W.E. (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*, 64-83, Cambridge: Cambridge University Press.

Nagy, W.E., Anderson, R. and Herman, P. (1987) Learning word meanings from context during normal reading, *American Educational Research Journal*, 24 (2), 237 - 270.

Nagy, W.E. and Herman, P.A. (1987) Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction, in McKeown, M.G. and Curtis, M.E. (eds.) *The nature of vocabulary acquisition*, 19-35, Hillsdale, NJ: Lawrence Erlbaum.

Nagy, W.E., Herman, P. and Anderson, R. (1985) Learning words from context, *Reading Research Quarterly*, 20 (2), 233-253.

Nassaji, H. (2003) Higher-level and lower-level text processing skills in advanced ESL reading comprehension, *The Modern Language Journal*, 87 (2), 261-276.

Nassaji, H. (2007) Schema Theory and Knowledge-Based Processes in Second Language Reading Comprehension: A Need for Alternative Perspectives, *Language Learning*, 57 (1), 79-113.

Nation, I.S.P. (1990) *Teaching and Learning Vocabulary*, Boston: Heinle and Heinle.

Nation, I.S.P. (1997) The language learning benefits of extensive reading, *The Language Teacher*, 21 (5), 13-16.

Nation, I.S.P. (2001) *Learning Vocabulary in Another Language*, Cambridge: Cambridge University Press.

Nation, I.S.P. (2006) How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63 (1), 59-82.

- Nation, P. (2007) The four strands, *Innovation in Language Learning and Teaching*, 1 (1), 2-13.
- Nation, P. and Beglar, D. (2007) A vocabulary size test, *The Language Teacher*, 31 (7), 9 - 13.
- Nation, P. and Wang, K.M.T. (1999) Graded readers and vocabulary, *Reading in a Foreign Language*, 12 (2), 355-380.
- Nation, P. and Waring, R. (1997) Vocabulary size, text coverage, and word lists, in Schmitt, N. and McCarthy, M. (eds.) *Vocabulary: Description, acquisition and pedagogy*, 6-19, Cambridge: Cambridge University Press.
- Nettles, D.H. (2006) *Comprehensive Literacy Instruction in Today's Classrooms: The Whole, the Parts, and the Heart*, California University of Pennsylvania, (http://www.ablongman.com/html/productinfo/nettles1e/contents/0205344259_9.pdf) Online, accessed 3 October 2009.
- Ng, S.M. (2001) The Brunei reading and language acquisition project, *International Journal of Educational Research*, 35 (2), 169-179.
- Ng, S.M., and Sullivan, C. (2001) The Singapore reading and English acquisition program, *International Journal of Educational Research*, 35 (2), 157-167.
- Nishizawa, H., Yoshioka, T. and Fukada, M. (2010) The impact of a 4-year extensive reading program, in Stoke, A. M. (ed.) *JALT2009 Conference Proceedings*. Tokyo: JALT. (<http://jalt-publications.org/archive/proceedings/2009/E035.pdf>) Online, accessed 6 December 2010.
- Norris, J. M. and Ortega, L. (2000) Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis, *Language Learning*, 50 (3), 417-528.
- Nunan, D. (1991) *Language Teaching Methodology: A Textbook for Teachers*, Hertfordshire, UK: Prentice Hall.
- Nunan, D. (1997) Developing standards for teacher-research in TESOL, *TESOL Quarterly*, 31 (2), 365-367.
- Nuttall, C. (1996) *Teaching Reading Skills in a Foreign Language*, Oxford, UK: Macmillan.

O'Brien, S. (2011) Building readers, how families can help children become better readers, The Parent Institute :NIS, Inc., (www.parent-institute.com, <http://falcon.sfcc.pvt.k12.wv.us/files/Building%20Readers%20March%202011.pdf>) Online, accessed February 23 2008.

Pallant, J. (2007) *SPSS Survival Manual (3rd edn.)* Open University Press: McGraw-Hill Education.

Pan, L. (1990) Sub-skills Approach and Extensive Approach to Reading in TEFL, *Journal of Cambridge Studies*, 4 (3), 112-120.

Panel on Education (papers) (1997) *Information paper, medium of instruction guidance for secondary schools*, (<http://www.legco.gov.hk/yr97-98/english/panels/ed/papers/ed1508-6.htm>) Online, accessed 9 December 2009.

Pang, E.S., Muaka, A., Bernhardt, E.B. and Kamil, M.L. (2003) *Teaching Reading*, International Academy of Education, IEA: Geneva available at (<http://www.ibe.unesco.org/International/Publications/pubhome.htm>) Online, accessed May 23 2008.

Paradis, M. (2000) Competence linguistique et connaissance metalinguistique en fonction de différentes formes d'enseignement: apport de neurosciences, in 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), 567-587.

Paribakht, T.S. and Wesche, M. (1997) Vocabulary enhancement activities and reading for meaning in second language vocabulary development, in Coady, J. and Huckin, T. (eds.) *Second language vocabulary acquisition: A rationale for pedagogy*, 174 – 200, Cambridge: Cambridge University Press.

Paribakht, T.S. and Wesche, M. (1999) Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing, *Studies in Second Language Acquisition*, 21, 195-224.

Pierson, H.D. (1988) The role of cultural and intellectual content in developing reading habits in a bi-lingual/multi-lingual setting, in Bickley, V. (ed) *Languages in education in a bi-lingual or multi-lingual setting*, , 347-357, H.K. Government Ed. Dept.

Pilgreen, J.L. (2000) *The SSR handbook: How to Organize and Manage a Sustained Silent Reading Program*, Boynton/Cook: Heinemann.

Pitts, M., White, H. and Krashen, S.D. (1989) Acquiring second language vocabulary through reading. A replication of the Clockwork Orange study using second language acquirers, *Reading in a Foreign Language*, 5 (2), 271-275.

Preator, S. (2010) Summer reading: Helping young readers find a “just right” book, (<http://www.imaginelearning.com/blog/index.php/2010/06/summer-reading-helping-young-readers-find-a-just-right-book/>) Online, accessed 25 October 2010.

Prowse, P. (2002) Top ten principles for teaching extensive reading: a response, *Reading in a Foreign Language*, 14 (2), 142-5.

Qian, D.D. (1999) Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension, *The Canadian Modern Language Review*, 26 (2), 282-308.

Qian, D.D. (2002) Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective, *Language Learning*, 52 (3), 513-536.

Reynanda W. and Jacobs, G. (2002) Extensive reading: Why aren't we all doing it?, in Richards, J. and Reynanda, W. (eds.) *Methodology in language teaching: An anthology of current practice*, 295-302, New York: Cambridge University Press.

Robb, T.N. and Susser, B. (1989) Extensive reading vs. skills building in an EFL context, *Reading in a Foreign Language*, 5 (2), 239-251.

Roszell, R. (2007) *Combining Extensive Reading and Intensive Vocabulary Study in a Japanese University*, unpublished doctoral dissertation, Temple University Japan, Tokyo.

Rott, S. (1999) The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading, *Studies in Second Language Acquisition*, 21 (4), 589-619,

Saragi, T., Nation, P. and Meister, G. (1978) Vocabulary learning and reading, *System*, 6 (2), 72-80.

Schmidt, F and Hunter, J. (2002) Are there benefits from NHST?, *American Psychologist*, 57 (1), 65-66.

Schmitt, N. (2000) *Vocabulary in Language Teaching*, New York: Cambridge University Press.

Schmitt, N. (2008) Instructed second language vocabulary learning, *Language Teaching Research*, 12 (3), 329-363.

Schmidt, R. (1994) Deconstructing consciousness in search of useful definitions for applied linguistics, in Hulstijn, J.H., and Schmidt, R. (eds.) *Consciousness and second language learning: Conceptual, methodological and practical issues in language learning and teaching*, *Thematic issue of AILA Review - Revue de l'AILA*, 11, 11-26.

Schmidt, R.W. (1990) The role of consciousness in second language learning, *Applied Linguistics*, 11 (2), 129-158.

Scott-Bucleuch, R.L. (2005) *The Promise*, Oxford: MacMillan Heinemann ELT.

Sharwood-Smith, M. (1993) Input enhancement in instructed SLA, *Studies in Second Language Acquisition*, 15 (2), 165-79.

Sheu, S.P.H. (2003) Extensive reading with EFL learners at beginning level, *TESL Reporter*, 36 (2), 8-26.

Sheu, S.P.H. (2004) Students' reflections on the physical features of EFL graded readers, *TESL Reporter*, 37 (1), 18-33.

Shokrpour, N. and Gibbons, J. (1998) Register complexity and the inadequacy of readability formulas as a measure of difficulty, *IJOAL*, 24 (2), 19-36.

Shu, H., Anderson R.C. and Zhang, H. (1995) Incidental Learning of Word Meanings While Reading: A Chinese and American Cross-Cultural Study, *Reading Research Quarterly*, 30 (1), 76-95.

Singleton, D. (1999) *Exploring the Second Language Mental Lexicon*, Cambridge: Cambridge University Press.

Smith, K. (2006) A Comparison of “Pure” Extensive Reading with Intensive Reading and Extensive Reading with Supplementary Activities, *The International Journal of Foreign Language Teaching*, 12-16, Fall.

Smith, W. (2005) *Eye of the Tiger*, MacMillan Heinemann ELT: Oxford.

Sonbul, S. and Schmitt, N. (2009) Direct teaching of vocabulary after reading: is it worth the effort?, *ELT Journal*, 64 (3), 253-260.

Stahl, S. (1983) Differential word knowledge and reading comprehension, *Journal of Reading Behavior*, 15 (4), 33-50.

Stahl, S. (1999) *Vocabulary Development*, Cambridge, MA: Brookline Books.

Stahl, S. and Nagy, W. (2006) *Teaching Word Meanings*, Mahwah, NJ: L. Erlbaum.

Storey, P., Yu, V., Luk, J., Ma, A., Pang, M., Wong, C. and Wong, W. (1997) *Feasibility Study on extending extensive reading in Chinese and in English to all levels of primary and secondary schooling*, Hong Kong: Hong Kong Institute of Education.

Susser, B. and Robb, T.N. (1990) EFL Extensive Reading Instruction: Research and Procedure, *JALT Journal*, 12 (2), 161-185.

Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output, in Gass, S. and Madden, C. (eds.) *Input in second language acquisition*, 235-253, Rowley, MA: Newbury House.

Swain, M. (1996) Discovering successful second language teaching strategies and practices: From programme evaluation to classroom experimentation, *Journal of Multilingual and Multicultural Development*, 17 (2-4), 89-113.

Swain, M. (2005) The output hypothesis: Theory and research, in Hinkel, E (ed.) *Handbook of Research in Second Language: Teaching and Learning*, 471 – 483, London: Laurence Erlbaum Associates.

Swain, M. and Lapkin, S. (1995) Problems in output and the cognitive processes they generate: A step towards second language learning, *Applied Linguistics*, 16 (3), 371-391.

Swanborn, M.S.L. and de Glopper, K. (1999) Incidental word learning while reading: A meta-analysis, *Review of Educational Research*, 69 (3), 261-285.

Tanaka, H. and Stapleton, P. (2007) Increasing reading input in Japanese high school EFL classrooms: An empirical study exploring the efficacy of extensive reading, *Reading Matrix*, 7 (1), 115-131,
(http://www.readingmatrix.com/articles/tanaka_stapleton/article.pdf) Online, accessed 26 May 2008.

Tekmen, E.A.F. and Daloğlu, A. (2006) An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency, *Foreign Language Annals*, 39 (2), 220-243.

Thompson, B. (2002) What future quantitative social science research could look like: confidence intervals for effect sizes, *Educational Researcher*, April, 25-32.

Truscott, J. (1998) Noticing in second language acquisition: A critical review, *Second Language Research*, 14 (2), 103-135.

Tsang, W.K. (1996) Comparing the effects of reading and writing on writing performance, *Applied Linguistics*, 17 (2), 210-233.

Tu, H.F. (2004) *Effects of Task-Induced Involvement on Incidental Vocabulary Learning in a Second Language*, unpublished Master's Thesis, National Tsing Hua University, available at (<http://www.hss.nthu.edu.tw/~fl/thesis/tesol/905258.pdf>)

Tup, F. and Shu, L. (1997) "First World - Third World", Two extensive reading programmes at secondary level, in Jacobs, G.M., Davis, C., and Renandya, W.A. (eds.) *Successful strategies for extensive reading*, 10-24, Singapore: SEAMEO Regional Language Centre.

Ulijn, J.M. and Strother, J.B. (1990) The effect of syntactic simplification on reading EST texts as L1 and L2, *Journal of Research in Reading*, 13 (1), 38-54.

Valentine, J.C. and Cooper, H. (2003) Effect Size Substantive Interpretation Guidelines: Issues in the Interpretation of Effect Sizes, Washington, DC: What Works Clearinghouse.

VanPatten, B. (1992) Processing Instruction: An update, *Language Learning*, 52 (4), 755-803.

Vicary, T. (2005) *Skyjack*, Oxford, Oxford University Press.

Wang, J.H.Y. and Guthrie, J.T. (2004) Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between U.S. and Chinese students, *Reading Research Quarterly*, 39 (2), 162-186.

Waring, R. (1999) *Tasks for assessing receptive and productive second language vocabulary*, unpublished paper for Doctor of Philosophy thesis, University of Wales, Swansea, (<http://www1.harenet.ne.jp/~waring/papers/phd/title.html>) Online, accessed May 23 2008.

Waring, R. (2001) Research in extensive reading, *Studies in Foreign Languages and Literature*, 25 (1), (<http://www1.harenet.ne.jp/~waring/papers/kiyo2001.html>) Online, accessed April 23 2007.

Waring, R. (2002) *Getting an extensive reading program going*, (<http://www1.harenet.ne.jp/~waring/papers/languageMagazine.doc>) Online, accessed 23 June 2007.

Waring, R. (2010) (27 June) *Extensive reading discussion forum*, <http://groups.yahoo.com/group/ExtensiveReading>.

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), 130-163.

Watanabe, Y. (1997) Input, intake and retention: Effects of increased processing on incidental learning of foreign language vocabulary, *Studies in Second Language Acquisition*, 19 (3), 287-307.

Webb, S. (2009) The effects of receptive and productive learning of word pairs on vocabulary knowledge, *RELC Journal*, 40 (3), 360-376.

Wesche, M. and Paribakht, T.S. (1996) Assessing L2 vocabulary knowledge: Depth versus breadth, *The Canadian Modern Language Review*, 53 (1), 13-40.

Wesche, M. and Paribakht, T.S. (2000) Reading-based vocabulary exercises in second language vocabulary acquisition, *The Modern Language Journal*, 84 (2), 196-213.

West, M. (1932) *New method readers*, London: Longman, Std. ed.

Wigfield, A., Eccles, J.S., Yoon, K.S., Harold, R.D., Arberton, A., Freedman-Doan, C and Blumenfeld, P.C. (1997) Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study, *Journal of Educational Psychology*, 89 (3), 451-469.

Wilkins, D. (1972) *Linguistics in Language Teaching*, London: Arnold.

Williams, E. (1987) Classroom reading through activating content-based schemata, *Reading in a Foreign Language*, 4 (1), 1-7.

Williams, R. (1986) 'Top ten' principles for teaching, *ELT Journal*, 40 (1), 42-45.

Williams, E. (2007) Extensive reading in Malawi: inadequate implementation or inappropriate innovation? *Journal of Research in Reading*, 30 (1), 59-79.

Wong, C.K. (2001) What we know after a decade of Hong Kong extensive reading scheme, *Eric Report*, University of Connecticut, Eric Clearinghouse.

Wong L.S. (1993) *A study of extensive reading in secondary 1 English class*, unpublished paper for the Master of Education programme, University of Hong Kong, School of Education.

Yamazaki, A. (1996) *Vocabulary acquisition through extensive reading*, unpublished Ph.D. Dissertation, Temple University, Japan.

Yang, A. (2001) Reading and the non-academic learner: a mystery solved, *System*, 29 (4), 451-466.

Yang, M.N. (2007) Language Learning Strategies for Junior College Students in Taiwan, *The Asian EFL Journal*, 9 (2), 35-57.

Yu, F.Y. (1979) *Report on an Assessment of the Standard of English of Pupils in HK*, Working papers in language and language teaching, 1, HK: Language Centre, University of Hong Kong.

Yu, S.L. (2000) *Strategy use by good and poor Chinese ESL Readers in comprehending easy and difficult expository text*, unpublished Master of Arts thesis, Hong Kong University.

Yu, V.W.S (1993) Extensive reading programs – how can they best benefit the teaching and learning of English?, *TESL Reporter*, 26 (1), 1-2.

Zahar, R., Cobb, T. and Spada, N. (2001) Acquiring vocabulary through reading: Effects of frequency and contextual richness, *Canadian Modern Language Review*, 57 (4), 541-572.

APPENDIXES

The appendixes are contained on the CD ROM attached to this thesis in PDF format.

These files can be read by downloading a free Adobe Reader from:

<http://get.adobe.com/reader/>