

# **A META-ANALYSIS OF EXTENSIVE READING RESEARCH**

---

A Dissertation Proposal  
Submitted  
to the Temple University Graduate Board

---

In Partial Fulfillment  
of the Requirements for the Degree of  
Doctor of Education

---

By  
Takayuki Nakanishi  
January, 2014

## Examining Committee Members

David Beglar, Advisory Chair, Teaching and Learning  
Yo In'nami, External Member, Shibaura Institute of Technology  
Kazuya Saito, External Member, Waseda University  
Paul Nation, External Member, Victoria University of Wellington  
Rob Waring, External Member, Notre Dame Seishin University

©

Copyright

2014

by

Takayuki Nakanishi

## **ABSTRACT**

### A Meta-Analysis of Extensive Reading Research

Takayuki Nakanishi

Doctor of Education

Temple University, 2014

Major Advisors: Professors David Beglar, Yo In'nami, and Kazuya Saito.

The purpose of this study is to investigate the overall effectiveness of extensive reading, and whether learners' ages impact learning from extensive reading differently. The second purpose is to investigate whether the length of time that second language learners engaged in extensive reading influenced outcome measures, and if an effect is found, identify effective periods of time for learners to engage in extensive reading. Meta-analysis was used to investigate the trends shown by past empirical research, chart future research directions, and provide a means to create alternative hypotheses for future research. Two types of empirical studies were conducted: group contrasts of studies that included control groups and pre-post contrasts of studies that only include experimental groups. This meta-analysis included 34 studies that provided 43 unique effect sizes (22 effect sizes for group contrasts and 21 effect sizes for pre-post contrasts) and a total sample size of 3,942 participants. Students who received extensive reading instruction outperformed those who did not. The effect size was small for group contrasts ( $d = .46$ ) and medium for pre-post contrasts ( $d = .71$ ). Moderator variables for each contrast were examined to assess the impact of learners' ages and the length of instruction; however, all groups substantially overlapped with each other in terms of their confidence intervals,

indicating no statistically significant difference among the groups. There was a small effect size for university students for the group contrast ( $d = .48$ ), a medium effect for high school students ( $d = .61$ ), a large effect for university students ( $d = 1.12$ ), and a large effect for adults ( $d = 1.48$ ) for pre-post contrasts. In terms of the length of instruction, both one semester of instruction ( $d = .36$ ) and one year of instruction ( $d = .52$ ) produced a small effect for group contrasts, while one year of instruction produced a medium effect ( $d = .74$ ) for pre-post contrasts. In sum, the available extensive reading research to date suggests that extensive reading improves students' reading proficiency and should therefore be a part of foreign language reading curricula.

## **ACKNOWLEDGMENTS**

I would like to express my sincere gratitude to my supervisors David Beglar, Yo In'nami, and Kazuya Saito. I am deeply indebted to them for their valuable feedback and comments throughout the process of dissertation writing and for making it possible for me to produce a higher quality final manuscript. Dr. David Beglar patiently read my drafts and provided feedback, guidance, and support so that I could continue writing and complete the dissertation. Drs. Yo In'nami and Kazuya Saito gave me constructive feedback on the earlier versions of the manuscript. Without their hard work and willingness to trust me, this project would not have been possible. I would also like to thank Professor Paul Nation and Dr. Rob Waring for serving as members of my oral dissertation committee. Their advice also helped to improve the quality of this dissertation.

**To my parents and relatives**

Their support and encouragement  
throughout my education and beyond were  
of the utmost importance. I dedicate this work to them.

## TABLE OF CONTENTS

	PAGE
ABSTRACT .....	iv
ACKNOWLEDGMENTS .....	vi
DEDICATION .....	vii
LIST OF TABLES .....	xii
LIST OF FIGURES .....	xiii
CHAPTER	
1. INTRODUCTION .....	1
The Background of the Issue .....	1
Statement of the Problem.....	6
Purposes of the Study.....	9
Audience for the Study .....	12
Delimitations of the Study .....	13
Definitions of Terms.....	14
The Organization of this Study.....	15
2. REVIEW OF THE LITERATURE .....	16
Extensive Reading .....	16
Principles for Setting Up an Extensive Reading Program .....	21
Extensive Reading Research.....	27
Meta-Analytic Studies of Extensive Reading.....	33
Junior High School Contexts.....	34

High School Contexts.....	37
University Contexts.....	48
Studies with Adult Participants .....	71
Chapter Summary .....	75
3. RESEARCH SYNTHESIS .....	79
Meta-Analysis .....	79
Criticisms of Meta-Analysis.....	86
Meta-Analytic Studies in the Field of Second Language Acquisition.....	89
Meta-Analyses of Extensive Reading .....	96
Purpose of the Study .....	99
Research Questions .....	101
4. METHODS .....	102
Procedures.....	102
Searching the Literature .....	103
Criteria for Inclusion.....	105
Analyses.....	109
Calculation and Interpretation of Effect Sizes .....	109
Fixed-Effect Model and Random-Effect Model .....	112
Homogeneity of Variance Tests .....	113
Descriptive Statistics .....	114
Outliers .....	115
Coding the Primary Studies.....	121
Rationale for Each Moderator .....	123

Publication Bias (Sensitivity Analysis) .....	125
Data Summary .....	130
5. RESULTS .....	134
The Research Synthesis .....	134
Publication Dates.....	134
Research Features.....	136
Meta-Analysis .....	138
Moderator Variable Analysis .....	141
6. DISCUSSION.....	149
Overall Effectiveness of Extensive Reading.....	149
Learners' Age.....	150
Length of Instruction.....	153
Outcome Measures.....	155
Deficiencies in Extensive Reading Research.....	157
Comparison with Other Meta-Analyses of Extensive Reading .....	159
7. CONCLUSION.....	162
Summary of the Findings.....	162
Limitations .....	163
Implications for Research Methodology.....	164
Suggestions for Future Research .....	169
Final Conclusions.....	171
REFERENCES CITED.....	172
APPENDICES	

A. SUMMARY OF EXTENSIVE READING STUDIES BY DAY AND BAMFORD.....	191
B. SUMMARY OF EXTENSIVE READING STUDIES BY HORST .....	192
C. SUMMARY OF EXTENSIVE READING STUDIES BY IWAHORI .....	193

## LIST OF TABLES

Table	Page
1. Summary of Extensive Reading Studies .....	78
2. Summary of Meta-Analytic Studies in Second Language Acquisition.....	95
3. Descriptive Statistics (Group Contrasts) .....	116
4. Descriptive Statistics (Pre-Post Contrasts).....	118
5. Effect Sizes Aggregated for Calculating the Overall Effect of Extensive Reading (Group Contrasts).....	120
6. Effect Sizes Aggregated for Calculating the Overall Effect of Extensive Reading (Pre-Post Contrasts) .....	121
7. Coding Table .....	126
8. Coded Variables Across Studies .....	127
9. Publication Dates of Studies Included in Appendices A, B, C, and in this Meta-Analysis .....	136
10. Research Features .....	139
11. Summary of the Interpretation of Effect Sizes .....	140
12. Overall Meta-Analysis for Group Contrasts .....	142
13. Overall Meta-Analysis for Pre-Post Contrasts .....	143
14. Participants' L1 and Tests Used in Each Study .....	147

## LIST OF FIGURES

Figure	Page
1. Funnel plot of group contrasts .....	131
2. Funnel plot of pre-post contrasts.....	132

## **CHAPTER 1**

### **INTRODUCTION**

#### **The Background of the Issue**

Starting from junior high school, formal English education in Japan continues for six years until students graduate from high school. Many Japanese students pursue further studies in college where English is typically a required part of the curriculum for the first two years. However, despite having eight years of formal English education, many Japanese students do not read a great deal of English, in part because they are not provided with sufficient opportunities to do so. Instead, a focus is placed on studying English grammar and translating English into Japanese because these skills are widely considered necessary for passing competitive Japanese university entrance examinations. Kikuchi (2006) investigated the types of questions that appear on Japanese university entrance examinations and reported that many questions require examinees to translate sentences from English to Japanese. Nakanishi (2008) stated that because of the presence of translation questions on the entrance examinations of public and private universities, high school teachers are pressured to teach students to translate English sentences into Japanese.

The emphasis on translation explains the widespread use of the grammar-translation method, and the use of this method means that most Japanese students have read a relatively small amount of English even after eight years of study. Reading is a crucial source of input for L2 development, especially in EFL (English as a foreign language) contexts such as Japan. Partly for this reason, extensive reading—which Davis (1995) defined as providing students with “the time, encouragement, and materials to read pleurably, at their own level, as many books as they can, without the pressures of testing or marks” (p. 329) —is gaining in popularity, as it is gradually being viewed as one of the most effective ways to increase the amount of comprehensible English input to students both inside and outside of the classroom (Takase, 2010, pp. 13-16).

Because of the inordinate amount of attention that many Japanese secondary school English teachers place on memorizing vocabulary and analyzing sentence structure, and because of a lack of opportunities to engage in authentic, communicative reading tasks, many Japanese students cannot adequately comprehend the meaning of written English texts despite a reasonably good understanding of the component parts (e.g., the lexis and morpho-syntax). Given this situation, extensive reading can plausibly play an important role in encouraging Japanese students to read

more, helping them comprehend English-language texts more fully, and leading them to enjoy reading in English.

There has been a growing interest in extensive reading as some authorities view it as the most promising way to improve students' reading proficiency.

Researchers have demonstrated the effectiveness of extensive reading in many contexts (e.g., Elley & Mangubhai, 1981, 1983; Greenberg, Rodrigo, Berry, Brinck, & Joseph, 2006; Hafiz & Tudor, 1990; Lai, 1993; Mason & Krashen, 1997; Robb & Susser, 1989; Williams, 2007), and have provided evidence supporting its positive effect on numerous cognitive and affective factors (e.g., Camiciottoli, 2001; Horst, 2005; Leung, 2002; Takase, 2004, 2007a; Yamashita, 2004, 2007). Implementing extensive reading has become easier because numerous classroom activities using graded readers have been described by a number of authors (e.g., Bamford & Day, 2004; Nakanishi, 2005). Bamford and Day (2004) stated, "the activities help teachers introduce extensive reading to students, organize and introduce suitable reading material, motivate and support, and monitor and evaluate reading" (p. 4). Depending on how the activities are incorporated into the foreign language curriculum, they can facilitate development in many language domains, including reading comprehension, strategy use, fluency development, vocabulary acquisition and consolidation, and the

acquisition of morpho-syntax. Nakanishi (2005), for instance, proposed a pair work activity in which students read a graded reader as homework, and then discuss the book in terms of issues such as the quality of the story and the plot. During the discussion, students were able to reinforce vocabulary they had met while reading and develop productive knowledge of the lexis by using and recycling it. The students were asked to repeat the activity with a new partner so that they could further practice using the vocabulary and listen to a different story. It was believed that this task also facilitated the development of the learners' aural comprehension skills.

Despite the number of articles published on the topic of extensive reading over the past two decades, two issues need to be investigated further. First, most extensive reading studies have been conducted independently of one another and they have generally been short-term studies taking place over one academic semester or one academic year. Consequently, it is difficult to determine the effect of extensive reading on reading rate, reading comprehension, and vocabulary acquisition given the varying results of these individual studies. Rephrasing this issue using meta-analytic terms, Valentine and Cooper (2003) stated, "Readers need to know if the intervention's effects are large or small, meaningful or trivial" (p. 1). The second issue concerns the lack of replication studies. Such studies are important because they

enable researchers to more precisely identify the effects of an intervention and the stability of those effects across different educational contexts. As the number of replications increases, the results can be considered more reliable, particularly if the findings are in fundamental agreement.

One approach to addressing these issues is to conduct a meta-analysis, which is “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (Glass, 1976, p. 3). Norris and Ortega (2006b) expressed the importance of meta-analysis as follows: “We envision synthetic methodologies as advancing our ability to produce new knowledge by carefully building upon, expanding, and transforming what has been accumulated over time on a given topic of interest for the research community” (p. 14). Once a sufficient number of studies on a particular topic have been conducted, a meta-analysis can confirm whether the combined effects of the treatments are important according to specific statistical criteria and provide greater insight and understanding than that provided by any single study because the *N*-size of a meta-analysis is far greater than that of any single study; this means that the conclusions drawn from a meta-analysis are also more valid and trustworthy. A meta-analysis also provides a summary of what has been accomplished in a particular field and thereby identifies

areas that have received a great deal of attention as well as those in need of further research. A further benefit is that the results of a meta-analysis can make researchers more aware of the characteristics of well-designed studies. Such designs can be replicated and included in future meta-analyses; furthermore, the meta-analysis itself can be replicated as has been done in meta-analyses of L2 instruction (Norris & Ortega, 2000; Doughty, 2003) and corrective feedback (Lyster & Saito, 2010).

The number of meta-analytic studies has increased rapidly in the past several decades because they allow researchers to arrive at relatively definitive conclusions concerning educational treatments. In 1976, meta-analyses appeared in only six studies, but by 1982 the number had increased to 120 (Rosenthal, 1987). White (1994) demonstrated the rapid increase of meta-analytic studies between 1974 and 1989 using three databases, PsycINFO, ERIC, and Social SciSearch, and Fields (2003) described the increase in the number of articles about or using meta-analysis published in science journals from 1981 to 2001.

### **Statement of the Problem**

One methodological gap in the field of second language acquisition is the lack of the use of meta-analysis. Although meta-analyses have been used frequently in fields such as medicine, relatively few researchers (e.g., Blok, 1999; Goldschneider &

DeKeyser, 2001; Masgoret & Gardner, 2003; Norris & Ortega, 2000, 2001; Ross, 1998; Sahari, 1997; see Norris & Ortega [2006a] for a collection of articles on meta-analysis in language learning and teaching) have conducted meta-analyses in the field of second language acquisition.

Using three databases, Academic Search Premier, ERIC, and MLA International Bibliography, I located 8,084 articles with *meta-analysis* in their titles as of March, 2010; however, the vast majority of the articles were unrelated to the field of second language acquisition. Next, *English* and *language* were added as keywords, and 1,154 articles were identified. When these 1,154 articles were examined manually, 23 articles were found to be related to the field of SLA. Only five of the 23 articles were published before 2000, a fact that shows that researchers have come to realize the value of conducting meta-analyses in the field of SLA relatively recently. Among the published studies, two small-scale meta-analytic studies (Krashen, 2007; Wang, 2010) were conducted on the topic of extensive reading. This lack of meta-analytic studies constitutes a major gap in the second language acquisition literature given the power of such studies to shed light on the effectiveness of particular pedagogical tasks such as extensive reading and to identify areas in need of further study.

The second problem is that although extensive reading researchers have repeatedly stated that extensive reading produces a multitude of positive effects, it is not clearly understood whether a learner's age (e.g., elementary school students, junior high school students, high school students, university students, and adults) impacts learning from extensive reading differently. It might be that extensive readings has differential effects on learners in different age groupings, or that age acts as a proxy for reading proficiency level, and that extensive reading has differential effects on learners at different proficiency levels.

The third problem concerns the effect of the length of time learners engage in extensive reading. Although there are suggestions in the extensive reading literature that extensive reading should be continued over a period of several years, it is not clear what the empirical evidence says regarding how long teachers should have students engage in extensive reading if they are to obtain particular results. It is also important to discover whether short-term extensive readings programs have beneficial effects, and if so, what those effects are.

The fourth problem concerns a lack of knowledge regarding how much students need to read to have a reasonable chance of obtaining a particular result. Second language reading authorities have proposed various reading goals, such as one

book per week (Day & Bamford, 1998), 500,000 running words per year (Nation, 2009), and one million words (Sakai & Kanda, 2005), but there is no empirical research to support the validity of these proposals.

Fifth, researchers lack precise knowledge of the measurements (e.g., TOEFL and IELTS examinations) utilized in previous studies. One by-product of a meta-analysis is to produce a count of dependent variables used in previous studies. This information is needed as it helps researchers decide what kinds of measurements should be administered to participants in future studies of extensive reading.

The final problem concerns avenues for future extensive reading research. The increase in the use of extensive reading has resulted in a greater number of studies of extensive reading; however, these studies cover a wide variety of topics, such as reading fluency development, reading comprehension improvements, and vocabulary acquisition. A more well coordinated approach to research might produce findings in areas that are currently not well understood.

### **Purposes of the Study**

The first purpose of this study is to investigate the overall effectiveness of extensive reading by aggregating the results of previous studies. Aggregating a large

number of individual studies provides more reliable results, a clearer overall picture of the results of the empirical research that has been conducted to date, and information regarding the types of research that have not been conducted in the past and that therefore might be conducted in the future. In order to accomplish this, a number of academic databases were used to identify published studies of extensive reading. Increasing the sample size results in more statistical power, which is one of the advantages of meta-analysis. An important contribution of the current study is its potential to highlight methodological problems, such as a lack of a control group and small *N*-sizes, in existing extensive reading studies.

The second purpose is to identify whether learners' ages impacts learning differently. Identifying when learners should begin extensive reading is an important pedagogical issue that benefits instructors and learners. Furthermore, information about age differences can help instructors and learners make appropriate short- and long-term plans for reaching various learning goals.

The third purpose is to investigate whether the length of time second language learners engage in extensive reading influences outcome measures. If an effect is found, the effects of different lengths of time will be investigated. Understanding the length of time that needs to be spent on extensive reading is beneficial for instructors

when creating objectives for reading courses and is a key issues in extensive reading pedagogy.

The fourth purpose is to investigate whether the amount read influences test scores, and what kinds of tests are utilized to determine the effects of extensive reading. If the amount read influences the reported outcomes, the results can provide guidelines concerning how much learners should read if they are to make significant gains. Furthermore, investigating the kinds of tests utilized in each study can help future researchers wishing to use the same measurements to replicate a study or compare their results with those obtained in other contexts.

The fifth purpose of this study is to suggest directions for extensive reading research. Summaries of the extensive reading research (e.g., Day & Bamford, 1998; Horst, 2005; Iwahori, 2008) have shown an overall picture of what aspects of extensive reading have been reasonably well investigated and what needs to be done to produce progress in the field. Based on the three studies, reading comprehension and reading speed are areas that researchers have investigated relatively frequently, while, the effects of extensive reading on the development of writing skills has been under-researched. This study can provide updated information concerning future research directions.

### **Audience for the Study**

The first audience for this study is researchers considering investigating extensive reading. They can benefit from the results of this study, as they shed light on the research that has been conducted to date and research areas that have not been sufficiently investigated. In this study, information about a large number of extensive reading studies is described in terms of issues such as sample size, treatment period, independent and dependent variables used, and educational outcomes.

This study is also valuable for researchers interested in meta-analysis. Meta-analysis is a useful analytical technique because it aggregates primary studies and compares the effect sizes obtained in each study. While the *N*-size in many single studies ranges from 20 to 30 participants, the total number of participants can be in the thousands when studies are combined. Thus, meta-analysis addresses the sample size problem that is all too common in the field of second language acquisition.

The third audience for this study is reading teachers. Teachers who are interested in extensive reading can acquire information about a large number of extensive reading studies and see the overall effects of extensive reading. This information is useful when they consider whether to implement extensive reading,

how to implement it, how much their students should read, and the probable outcomes of an extensive reading program.

### **Delimitations of the Study**

The findings of this study cannot be generalized to all contexts; however, with a moderator variable analysis, it is possible to understand contrasting effects in divergent circumstances, such as EFL and ESL contexts, as well as elementary school, secondary school, and university contexts. Researchers, however, can use the results from studies conducted in specific educational contexts. For example, instructors who are teaching extensive reading at a university for one semester can make use of the results of studies that were conducted in a university context similar to theirs. Most extensive reading research has been conducted in Asian contexts; thus, it is difficult to generalize those results to English language learners in regions such as Europe or South America given differences in native language (and language distance), culture, attitudes toward English, and opportunities to encounter English and speakers of English.

## Definitions of Terms

Definitions of the terms utilized in this study are provided in this section.

These terms are described in more detail in Chapter 2.

*Teaching reading* refers to combinations of reading strategies, vocabulary development, grammar accuracy, and reading fluency. It does not refer to grammar translation. To illustrate the definition of reading and what teaching reading entails, Grabe and Stoller (2011) stated, “the mastery of reading requires not only the integration of comprehension abilities but also the development of a very large vocabulary and a reasonably good command of grammar resources” (p. 129).

*Extensive reading* is defined in two ways in this study. The first, a pedagogical definition concerning how teachers can conduct extensive reading, has been used by a number of reading authorities (e.g., Davies, 1995; Day & Bamford, 1998). This definition refers to reading in quantity and providing learners with materials they can read enjoyably and that help them develop the skills needed to understand English texts. The second definition is a process-oriented definition that states that extensive reading is a way of processing text fluently and with high levels of comprehension. Some learners, particularly low reading proficiency learners, might not always be

reading extensively from a processing point of view, even when reading a text appropriate for their reading proficiency level.

### **The Organization of this Study**

Chapter 2, Review of the Literature, is divided into four main sections:

Extensive Reading, Principles for Setting Up an Extensive Reading Program,

Extensive Reading Research, and Meta-Analytic Studies of Extensive Reading.

Chapter 3, Research Synthesis, covers four topics: Meta-Analysis, Meta-Analytic

Studies in the Field of Second Language Acquisition, Purposes of the Study, and

Research Questions. In Chapter 4, Methods, I describe how the extensive reading

studies included in this study were identified, how the primary studies were coded,

and how effect sizes were determined. The results of the study are presented in

Chapter 5. In Chapter 6, Discussion, the results are discussed and pedagogical

implications are presented. A summary of the findings, the limitations of the study,

suggestions for future research, and final conclusions are provided in Chapter 7.

## **CHAPTER 2**

### **REVIEW OF THE LITERATURE**

#### **Extensive Reading**

For approximately 30 years, the grammar-translation method has been used to teach reading in Japan. In this method, students are trained to pay attention to each word and grammatical structure, and consult a dictionary every time an unknown word is encountered. Students are usually provided with a one- or two-page English text that is followed by comprehension questions. In many secondary school classrooms, it takes two or more classes to finish one short text, and by the time the students finish reading a text using this approach, they often do not remember what they have read and therefore have difficulty answering comprehension questions. Another problem with the grammar-translation method is that students do not read very much, even after years of study, as reading in this method proceeds extremely slowly. This problem has been noted by L2 reading specialists, such as Grabe and Stoller (2011), who stated that “most L2 readers are simply not exposed to enough L2 print (through reading) to develop fluent processing....Nor do they have enough exposure to build a large recognition vocabulary” (p. 50).

An alternative method of teaching reading known as extensive reading (ER) has drawn a great deal of attention in the past two decades. Extensive reading is an approach to teaching reading that provides large quantities of easily comprehensible English books to students, thereby improving their reading proficiency, while allowing them to enjoy the learning process (Day & Bamford, 1998). This approach differs completely from the traditional grammar-translation approach in terms of the quantity of reading, the difficulty level of the reading, the degree of autonomy enjoyed by the students, and the motivation to continue reading.

Extensive reading has been found to provide numerous language learning benefits. For instance, Day and Bamford proposed that one of the purposes of extensive reading is to increase students' enjoyment of reading; thus, they recognized the importance of affect in reading development. Davis (1995) emphasized the importance of incorporating extensive reading into foreign language curricula by stating that any ESL, EFL, or L1 classroom is the poorer for the lack of an extensive reading program, and teachers are unable to promote pupils' language development as effectively if such a program is not present (p. 329). In a similar vein, Grabe and Stoller (2011) placed high importance on reading in academic contexts, in which they stated "reading provides a major source of input for further students learning of both

language and content information. Furthermore, reading can generate increased interest and motivate students to explore topics further through additional reading” (p. 129). Moreover, extensive reading also benefits the development of reading strategies. Burrows (2012) investigated the relationship between extensive reading and reading strategy development in terms of change in reading comprehension and concluded that the combination of extensive reading and reading strategy instruction enhances L2 reading comprehension.

Research on extensive reading has primarily focused on the development of vocabulary knowledge. Previous L2 vocabulary researchers have established a rough estimate in terms of how many words readers need to know (e.g., 95-98% coverage [Nation, 2001]) to be able to read a text independently. Nation and Wang (1999) suggested that learners need to read one graded reader every one or two weeks to benefit from extensive reading. For example, although previous literature suggests that 8-10 reading exposures can give learners a reasonable chance of acquiring an initial receptive knowledge of words, Laufer (2005) pointed out that “In order for words to be met 10 times in reading, learners would need to read 1-2 graded readers per week. The typical learner simply does not read this much” (p. 241).

Nation (2009) argued for the effectiveness of extensive reading on vocabulary acquisition, especially through the use of simplified texts such as graded readers, instead of using authentic materials. Nation compared an original version of *Dracula* with the simplified version and found that the vocabulary was controlled in such a way that there were very few low frequency words and the high frequency words were repeated numerous times in the text. This type of repetition is important if learners are to acquire form-meaning mappings. Moreover, given the importance of the 2,000 high frequency words of English, acquiring and automatizing these words is extremely important, as they provide a base for the skilled reading of virtually all English texts. Nation also discussed the effectiveness of extensive reading where the development of reading fluency is concerned. Fluency increases are important, as many second language readers of English read well below the 200 word-per-minute rate identified as a minimum for some types of skilled reading. In this vein, a strong relationship between reading speed and reading comprehension has been found in both first and second language studies of reading (e.g., Beglar, Hunt, & Kite, 2012; Nuttall, 2005). Learners who enjoy reading read faster, read more, and eventually comprehend better. Nuttall calls this the “virtuous circle of the good reader” (p. 127).

Researchers (Day & Bamford, 1998; Grabe, 2009) have also found no evidence in favor of authentic texts over graded readers where foreign language acquisition is concerned. Accordingly, reading texts should be selected based on the learners' current vocabulary levels, as books that are largely incomprehensible make weak contributions to acquisition. In this vein, Schmitt (2008) stated that the vocabulary load can be "fine-tuned for the learner's level, and systematically recycled" (p. 349). Schmitt argued that students need to have an explicit vocabulary learning component, but that extensive reading can increase the number of exposures of intentionally learned vocabulary. Numerous researchers (e.g., Anderson, 2009; Grabe, 2009; Grabe, & Stoller, 2011; Nuttall, 2005) also support the notion that extensive reading can enhance learners' knowledge of the contextual usage of L2 lexis. As noted by the above reading authorities, the key to success in foreign language vocabulary acquisition is found in repeated opportunities to meet the same lexis repeatedly in communicative contexts; extensive reading can be of enormous assistance in this endeavor. Extensive reading is also potentially more valuable for young adults and adults than younger learners because older learners have more opportunities to meet the same vocabulary, as they can read more books in the same amount of time compared with children.

## **Principles for Setting Up an Extensive Reading Program**

Elley (1986, p. 212) made five suggestions regarding setting up and running extensive reading programs or implementing extensive reading for children. To the best of my knowledge, Elley's suggestions formed the first set of principles for creating an extensive reading program:

1. Provide massive doses of reading at an early stage, as this can help ESL children's language development.
2. Implement the Shared Book approach.
3. Ensure that teachers adhere to the program guidelines.
4. Use universal favorites, such as Cinderella, rather than books on local settings and events.
5. Understand that extensive reading programs are effective because they create high motivation to read, the emphasis is on meaning, and students are exposed to large amounts of high quality language.

Elley's list offers specific suggestions (see principles 4 and 5) regarding what materials to use and the importance of the teacher's understanding of extensive reading.

More than a decade later, Day and Bamford (1998) proposed a more detailed list of ten principles upon which successful extensive reading programs are based. These principles continue to attract considerable attention from extensive reading researchers:

1. Students read as much as possible.
2. A variety of materials on a wide range of topics is available.
3. Students select what they want to read.
4. The purposes of the reading are usually related to pleasure, information, and general understanding.
5. Reading is its own reward.
6. Reading materials are well within the linguistic competence of the students.
7. Reading is individual and silent.
8. Reading speed is usually faster rather than slower.
9. Teachers orient students to the goals of the program, explain the methodology, keep track of what each student reads, and guide students in getting the most out of the program.
10. The teacher is the role model of a reader for students.

Day and Bamford (2002) later published this list as the top ten principles for teaching extensive reading. These ten principles can also be seen as advantages of extensive reading over the traditional grammar-translation approach to teaching reading that is still used widely in Japan. The contribution that these lists bring to the field of second language reading is that students should read far more than they typically do in terms of quantity. The students also choose books to read, a form of autonomy that might lead them to enjoy what they read and therefore to read more. The last principle concerning the teacher as a role model has not been investigated, so it is not clear to what degree it is necessary or important, particularly given the fact that many students engaged in extensive reading read outside of class. Day and Bamford further suggested that teachers utilize the principles to reflect on their own teaching and their beliefs about teaching reading in general.

Grabe (2009) also listed 13 recommendations for engaging students in extensive reading (p. 327). These recommendations overlap with many of the principles mentioned above, but provide specific issues to be incorporated in class, such as allowing students to (a) read magazines and comic books in class, (b) talk about what they have read, (c) explain why the texts were (not) interesting to them, and (d) recommend reading texts to other students. The suggestions and principles by

the three authors provide a solid foundation for developing and maintaining extensive reading programs.

The ideas also share certain core elements concerning extensive reading and form three categories: the students' perspective, the teachers' perspective, and the extensive reading environment. These are summarized below.

The students' perspective includes:

1. Read as much as possible.
2. Choose their own books.
3. Read for pleasure
4. Read faster rather than slower.

The teacher's perspective includes:

1. Make sure that teachers understand the program and follow guidelines.
2. Guide students along the reading process.
3. Need to read to be the role model for students.
4. Learn what students like to read.
5. Tell students what you like to read and why.
6. Have students discuss what they read and recommend materials.

The extensive reading environment is summarized in the following ideas:

1. A wide range of books, including comics and magazines, are available.
2. Materials that are worldwide are available rather than texts from the local setting.
3. Materials at a wide variety of difficulty levels are available.
4. Time to read inside and outside of class is available.

Extensive reading, however, is viewed favorably by all writers. For instance, some have argued that the teachers do not appear to be teaching. Prowse (2002) noted that “a problem encountered in some cultures though is that 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” (p. 144). Grabe (2009) pointed out the same issue and suggested that it is the reason for the unpopularity of extensive reading in some schools. In response to this criticism, Takase (2007b) and Muto (2006) stated that teachers in an extensive reading program assume a very different role compared with the traditional teacher-fronted classroom. The teacher’s understanding of the program, thus, is one key to its success.

Nakanishi and Ueda (2011) specifically asserted that the teacher's role in extensive reading programs is as follows:

The teacher walked around the classroom, helping students choose books, consulting, advising students on reading or having short discussions about their books, sharing and exchanging impressions and sentiments—all on an individual basis. When students approached the displayed books to choose new ones, the teacher was there and talked a little about each of the books they examined. It was very important that the teacher knew something about (or, preferably, had read) most or all of the books in order to give appropriate, funny, and motivating comments to each student. This is a very important role for the teacher in an ER class (p. 3).

They stated that the role of the teacher is not to “teach” per se but rather to facilitate the students' learning on an individual basis and establish and maintain the conditions of a reading environment that encourages extensive reading (p. 4).

Along with the principles described above, extensive reading also differs from other forms of reading instruction in terms of the way assessment is conducted.

Teachers typically must award grades to students; therefore, criterion-referenced tests are often administered to assess whether students have learned what they were taught.

With extensive reading, on the other hand, teachers do not have to assess students on what they read. Instead, as shown in the next section, teachers or researchers can assess general reading comprehension, reading speed, or vocabulary improvement.

## Extensive Reading Research

In this section, I review studies that exemplify extensive reading research.

Next, I review studies that are particularly relevant to this study, as they are included in the meta-analysis presented in Chapter 4. The studies are divided into four learner age groups, which is a first moderator variable analyzed in the Results chapter of this study.

From the 1980's, one of the main issues extensive reading researchers examined was the relationship between reading and writing. Janopoulos (1986) investigated whether a positive correlation exists between L1 and/or L2 pleasure reading and L2 writing proficiency with 79 foreign graduate students. The participants produced a 1-hour writing sample and later completed a questionnaire asking the amount of time they spent reading each week for pleasure in the L1 and L2. The results showed a strong positive correlation ( $r = .76$ ) between the amount of L2 pleasure reading completed and L2 writing proficiency; however, no significant correlation between L1 reading and L2 writing proficiency was found ( $p = .07$ ). The author concluded that students who engage in more pleasure reading in the L2 are likely to be better writers. This is a reasonable conclusion because large amounts of reading should allow learners to deliberately and incidentally acquire knowledge of

L2 spelling patterns, lexis, and morpho-syntax. This knowledge should eventually be transferred to writing tasks.

Following Janopoulos, several researchers (Hafiz & Tudor, 1989; Lai, 1993; Mason & Krashen, 1997) attempted to test the hypothesis that exposure to large amounts of reading results in greater writing proficiency.

Hafiz and Tudor (1989) investigated the effects of extensive reading on students' reading and writing skills using graded readers with Pakistani children (aged 10-11) studying English. The experimental group ( $n = 16$ ) voluntarily joined a three-month extensive reading program, while the children in two control groups ( $n = 15$  each) received no special extensive reading treatment. Although the authors did not specify the number of words or books the students read, the program was conducted after school for one hour, five days a week for twelve weeks; therefore, the students should have read a relatively large amount of text. The participants completed the NFER (National Foundation for Educational Research) proficiency tests as pre- and posttests. The test consists of several reading and writing sections. For example, the reading section has questions such as identifying a synonym for an underlined word and in the writing section, participants write a sentence describing a picture. The results of paired  $t$ -tests indicated that the scores of the experimental

group increased statistically significantly on both the reading and writing measures. No statistically significant differences were found on reading for the control groups, but one of the control groups scored significantly higher on two parts of the writing test. As one of the limitations, the authors stated that it might have been due to the fact that the students remembered the content or the form of the tests from the pre-testing session in the first week. The authors stated that their finding supported Krashen's input hypothesis, indicating that extensive L2 input in an anxiety-free environment can contribute significantly to the enhancement of learners' receptive and productive language skills (p. 10).

Students' attitudes toward extensive reading have also been investigated. Camiciottoli (2001) investigated the factors influencing reading frequency and attitudes toward extensive reading by administering a 22-item questionnaire to 182 Italian college students. To investigate reading frequency, the dependent variables were item 13 (frequency of reading any type of English materials) and item 15 (number of English books read in the last 12 months). Item 17 (attitude based on reasons cited for limited reading) and item 19 (opinion expressed on the importance of reading) represented the dependent variables for attitude. The independent variables included sex, age, private lessons, past access to English books, self-rating

of reading ability, time spent in an English-speaking country, and the number of books read in Italian in the last 12 months. With item 13 as a dependent variable, the results of a multiple regression analysis indicated that sex,  $B = .15$ , time spent in an English-speaking country,  $B = .18$ , and the number of books read in Italian in the last 12 months,  $B = .27$ , significantly predicted L2 reading,  $R^2 = .19$ . The author found that the participants read relatively little English, but they nonetheless held positive attitudes toward reading. The results also showed that the participants did not read a great deal in their L1; thus, the author speculated that the lack of reading in the L1 was one reason for the lack of reading in English. The author suggested that the cause of the positive attitudes toward reading was the growing influence of English in the Italian language via English loanwords.

Camiciottoli's findings were supported by Yamashita (2004), who investigated the relationship between L1 and L2 reading attitudes, measured with a five-point Likert scale questionnaire, and students' performance in L2 extensive reading, which was operationalized as the number of pages each student read. The participants were 59 Japanese university students. The reading attitude questionnaire was administered in the first extensive reading class. A factor analysis was conducted to identify factors formed by the questionnaire items. Eight factors emerged from the

analysis and Pearson correlation coefficients were computed among the eight variables. Negative correlations ranging from  $-.30$  to  $-.45$  were found for Comfort, which indicated a positive feeling toward reading, and Anxiety, which indicated negative feelings toward reading in both Japanese and English. The most important correlations (ranging from  $.32$  to  $.63$ ) were between the same L1 and L2 attitude variables; thus, L1 and L2 reading attitudes appeared to share some commonalities. Yamashita claimed that there might be transfer from L1 to L2 reading in the affective domain (e.g., attitude toward reading). This finding added to the reports of previous researchers (Bernhardt & Kamil, 1995; Carrell, 1991; Perkins, Brutton, & Pohlmann, 1989; Yamashita, 2002) that transfer also occurs in the cognitive domain (e.g., skilled L1 readers read well in L2). Yamashita also argued that her finding that affective transfer occurs was important for the advancement of research into the relationship between L1 and L2 reading because it addressed a gap in the extensive reading literature.

However, contrasting findings have also been reported. Investigating what factors motivate Japanese high school students to read English extensively, Takase (2004, 2007a) reported that good L1 reading habits do not always transfer to L2 reading, stating that some students in her study (2007a) “could not abandon the

enjoyment that they experienced when reading in Japanese and shift to the effortful and less enjoyable experience of reading in English” (p. 12). Many second language learners are able to read books enjoyably and comfortably in their L1, but once they start reading books in the L2, they experience significant decreases in reading speed, and encounter many unknown words that lead them to rely on dictionaries. These factors can lead to a less enjoyable reading experience. These contrasting findings were interesting although more research is needed in this area.

Waring and Takaki (2003) investigated how much vocabulary was learned and retained after 15 Japanese university students read the graded reader version of *A Little Princess*. The researchers replaced 25 words in the book with non-words. The learning of the 25 non-words was investigated through a word recognition test, a multiple-choice test, and a meaning translation test. The three tests were administered three times: immediately after reading, one week later, and three months later. The findings showed that the three tests scores decreased substantially on every subsequent administration, a finding that indicated that students could learn new vocabulary through reading to a degree, but the knowledge was not retained well. Like Nation and Wang (1999), Waring and Takaki concluded that a large amount of reading is required to consolidate words that are learned but still fragile.

Leung (2002) conducted a qualitative study of extensive reading focusing on her own learning of Japanese over a 20-week period. Data were gathered from a learner diary, audio-recordings from private tutorial sessions, and vocabulary tests. This study is unique in that the process of extensive reading is described from a learner's perspective. Because it was a self-study, Leung faced a number of challenges at the beginning of the study, including the difficulty of finding appropriate level reading materials, disciplining herself to read regularly, and allocating time to read extensively. However, once these challenges were overcome, the results showed that extensive reading improved her vocabulary, reading comprehension, and created a positive attitude toward reading. She read 32 books and 1,260 pages of simple Japanese books. The biggest obstacle was finding appropriate level reading materials; thus, she concluded that the success of extensive reading partly depends on access to a large collection of appropriate reading materials.

### **Meta-Analytic Studies of Extensive Reading**

A crucial pedagogical issue concerns how early extensive reading should be started. For this reason, it is important to investigate possible differences among age groups. Thus, in the following sections extensive reading studies conducted with

junior high school, high school, and university students, and adult learners are reviewed. Age group is also the first moderator variable investigated in this study.

The studies in each section are reviewed in chronological order.

### **Junior High School Context**

Tsang (1996) conducted a study in Hong Kong comparing the effects of extensive reading and frequent writing assignments on writing performance with 144 Cantonese-speaking students in grades 7-10. The participants were divided into three groups of 48 students each: a control group, an extensive reading group, and a writing group; thus, there were twelve groups (four grades with three groups in each grade) that engaged in the 24-week treatment. The extensive reading group was required to read eight graded readers and the writing group was given eight essay-writing tasks to complete. In the first and last class, the participants wrote about “My favorite person.” The essays were randomized and graded by two independent raters. Scores on three sets of essays produced by the three groups were the dependent variables, the three groups and four grades were the independent variables, and the pretest was a covariate. A MANCOVA was conducted, and the results indicated that there was no statistically significant difference among the three groups; however, based on the

posttest gains, the author argued that the extensive reading group demonstrated a significant gain in content and in language use on the writing assignment.

Sheu (2003) conducted a study in Taiwan designed to investigate the effects of extensive reading on reading fluency and reading comprehension using two types of reading materials, graded readers and books for native English-speaking children. Three classes of grade 2 Chinese students in Taiwan, aged 13-14, participated in the study for two semesters. The participants were in two experimental groups and a control group. Thirty-one participants read only graded readers (GR) and 34 participants read books for native English-speaking children (BNESC). The participants in the control group ( $n = 33$ ) reviewed parts of previous English lessons they had trouble following, memorized vocabulary and grammatical rules, and completed textbook exercises, which included gap-filling, multiple-choice, and true/false questions. The participants were tested with six matching vocabulary questions, seven multiple-choice grammar questions, and seven true-false reading comprehension questions. Paired sample *t*-tests were conducted with the pre- and posttest scores. The results indicated that the GR group outperformed the other groups on all three question types, whereas the BNESC group achieved significant gains on grammar and reading comprehension, but not on vocabulary. The author

proposed that the recycling of vocabulary in the GR condition might have contributed to the differences in scores.

Matsui and Noro (2010) also investigated the effects of 10-minute sustained silent reading (SSR) at the beginning of class once a week on junior high school student's reading comprehension and reading speed over 10 months. Sixty-two students were in the control group and 60 students were in the experimental group. The experimental group followed the same procedure as that used in the previous study (Fujita & Noro, 2009), while the control group engaged in the regular class instruction instead of sustained silent reading. The students in the experimental group read from 3,440 words to 234,004 words ( $M = 18,907$  words) over the ten-month period. Reading speed and comprehension were measured using the EPER as pre- and posttests. The dependent variables were reading speed and reading comprehension scores and the independent variable was group (two levels). The results of 2 x 2 mixed ANOVA showed that there was a statistically significant difference between pre- and posttests on reading comprehension; however, there was no difference between the experimental group and the control group ( $p = .40$ ) on reading comprehension. On the other hand, the results for reading speed indicated that the experimental group significantly outperformed the control group ( $p = .001$ ),

as the control group increased from 89.29 to 94.79 wpm and the experimental group improved from 88.57 to 108.32 wpm. The authors noted that extensive reading requires “individually-targeted teaching” (p. 79), meaning that teachers need to be aware of each student’s proficiency level and the topics they prefer so that appropriate suggestions can be given. Furthermore, they also argued that SSR effectively improves reading comprehension if students engage in it for a sufficiently long period of time.

Only three studies conducted in a junior high school context show the difficulty of conducting extensive reading studies in that context. There is no study dealing with younger students in extensive reading research, which is a major gap in the field. It is also interesting to note that the three studies were conducted in different countries, which might indicate that the difficulty of conducting a study with younger participants faces researchers in many educational contexts. The issue is discussed in more detail in the Discussion Chapter.

### **High School Context**

Lai (1993) investigated the effect of a summer reading course on reading and writing skill development with 226 Chinese secondary school students. The participants were divided into three groups in terms of the amount of English study

they had completed before the study began. The three groups took part in a four-week 50-hour summer reading program, which provided approximately 45 hours of reading. The participants read graded readers. Group 1 ( $n = 126$ ) read an average of 16.0 books ( $SD = 7.5$ ), Group 2 ( $n = 88$ ) read an average of 18.5 books ( $SD = 7.4$ ), and Group 3 ( $n = 52$ ) read an average of 14.2 books ( $SD = 12.1$ ). A standardized reading test and a researcher-designed speed-reading test were administered to assess the learners' reading proficiency and reading speed. Paired  $t$ -tests of the pre- and posttest scores on the standardized reading tests indicated that the first two groups showed a significant improvement for both reading ability and reading speed. The third group, which was made up of less motivated students, did not make significant progress on the standardized reading test. Lai concluded that reading proficiency and reading speed improved more for students who read more books than those who read fewer books.

Fujimori (2006) investigated the effects of extensive reading on reading and listening comprehension with 240 first-year Japanese high school students. The control group was made up of second-year students who had not engaged in extensive reading in the previous year. The students in the experimental group read graded readers with approximately 900 headwords during the class for a total of

approximately 25,000 words read. The participants had one hour of extensive reading instruction per week over ten months in addition to a three-hour intensive reading and one-hour English drama performance classes. All the participants completed the pre-second-grade and third-grade reading and third-grade listening sections on the EIKEN STEP tests as a pretest and posttest. However, the control group did not take the pretest; therefore, only the posttest scores were compared. The results of two-way ANOVAs (independent variables were the pre-second-grade and third-grade reading and third-grade listening sections on the EIKEN STEP tests; the dependent variable was ten months extensive reading instruction) showed that the scores for the easier reading and listening comprehension items were better for the extensive reading group ( $M = 7.54$  and  $6.11$ , respectively) than for the control group ( $M = 6.92$  and  $5.30$ , respectively), which the author attributed to the use of  $i - 1$  level extensive reading materials. The author speculated that the many participants automatized their bottom-up processing abilities during the treatment.

Smith (2006) investigated whether extensive reading could be improved by writing post-reading reaction reports and whether that treatment was more effective than intensive reading instruction. The supplementary materials required students to write a brief summary of a book and recommendations to other classmates. The

participants were 15 and 16 year old high school students in Taiwan who were placed in three groups ( $N = 51$ ). The students in the first group engaged in intensive reading by reading short reading passages and answering comprehension questions. The second group engaged in extensive reading and wrote post-reading reaction reports. They submitted one reaction report every two weeks over one academic year. The third group was an extensive reading only group. These students spent each class reading, and they did not complete any writing activities afterwards. All the participants took the Edinburgh project on extensive reading (EPER) placement/progress tests as pre-and posttests. The results of independent samples  $t$ -tests indicated that the extensive reading only group outperformed both the intensive reading group ( $t = 2.28, p = .01$ ) and the extensive reading group with supplemental activity ( $t = 2.65, p = .00$ ). Smith concluded that extensive reading only was most effective and that reading large numbers of books is an important source of English reading proficiency gains.

Tanaka and Stapleton (2007) investigated the effect of extensive reading on reading comprehension and reading speed with 96 Japanese high school students who engaged in extensive reading for the first five to ten minutes of class over five months. The control group consisted of 94 students who had exactly the same English

instruction but who did not engage in extensive reading. The students read teacher-made materials, and they were encouraged to read graded readers if they felt confident enough to do so. Some students read graded readers outside of class as well. The students read 38 passages (6,505 running words) as well as an intermediate textbook they used in class that contained 4,224 running words. Reading comprehension was assessed with the third and fourth grades of the STEP tests as pre- and posttests. The authors used the number of items the students answered as an indicator of their reading speed. The posttests were compared using independent samples *t*-tests. The experimental group and control group differed significantly on the reading comprehension test ( $t = 2.50, p = .01$ ). The students ( $n = 18$ ) who read graded readers were taken out of the experimental group and the remaining 78 students were again compared to the same control group to examine the effect of graded readers on reading speed. The result indicated that there was no statistically significant difference among the groups ( $t = 1.31, p = .19$ ); therefore, the authors argued that the students who had read graded readers primarily accounted for the gains in the extensive reading group. Reading speed was compared the same way, but the *t*-test results were not significant. The authors concluded that English teachers in

Japan should consider increasing reading input using extensive reading both inside and outside of the classroom.

Takase (2007a) administered a questionnaire to 216 Japanese high school students to investigate their motivation and attitudes toward reading in Japanese and English. The participants also completed the reading section of the Secondary Level English Proficiency (SLEP) test as a pre- and posttest ( $M = 35.07$ ,  $SD = 5.70$  and  $M = 41.74$ ,  $SD = 6.56$ , respectively). The participants read from 0 to 311,142 words in English ( $M = 71,653$ ). Based on a principal components analysis and a multiple regression analysis with the factor scores of the questionnaire acting as predictors of the participants' motivation to read English books and the amount of reading in English as the dependent variable, she reported that Intrinsic Motivation for L2 Reading and Intrinsic Motivation for L1 Reading were the most influential motivational factors. These findings were similar to those of Mori (1999, 2002), who characterized the factors as Positive Intrinsic Value and Intrinsic Value of Reading. However, Takase argued that her study differed from Mori's in three ways. First, her participants were high school students whereas Mori's participants were university students who did not have to face a university entrance examination, which plays an important role in academic motivation in Japan. The other two aspects dealt with the

reading materials and the places where the students engaged in extensive reading.

The participants in Takase's study autonomously chose their books, while Mori's participants had fewer choices as they were directed to read SRA materials in the library.

Iwahori (2008) examined the effectiveness of extensive reading on reading rates and reading proficiency with 33 second-year Japanese high school students who engaged in seven weeks of extensive reading instruction. On average, the students read four books during the treatment period. A short story was used to measure the participants' reading rates and a 100-item C-test was used as a reading proficiency measure. No control group was included in this study. The two dependent variables were reading rate and C-test scores and the independent variable was the seven-week extensive reading treatment. Two paired *t*-tests were statistically significant ( $p < .0005$  for reading rate;  $p = .003$  for C-test). Iwahori concluded that extensive reading improves the reading rates and general language proficiency of Japanese high school students even when they read relatively few books in a short treatment period (p. 83). However, the results need to be interpreted cautiously because the difference between the pre- and posttest scores for the C-test was less than four points.

Cha (2009) conducted a study with ten high school participants who engaged in extensive reading and another ten had no treatment for 12 weeks (approximately one semester) in a Korean vocational high school. The experimental group was encouraged to read at least one graded reader each week while the control group did not read or receive any instruction during the period because the study was conducted during winter vacation. It turned out, however, that the experimental group read only an average of 4.3 books (range = 2-7 books) during the 12-week program. The participants took an L2 reading comprehension test designed to measure their reading rates and reading proficiency, and a vocabulary test as a pre- and posttest. A repeated-measures ANOVA indicated that the experimental group achieved a significant reading rate gain ( $p = .00$ ), but made non-significant gains on reading comprehension ( $p = .06$ ) and vocabulary ( $p = .13$ ); however, the reading comprehension and vocabulary posttest scores for the experimental group were far better than those of the control group. Cha suggested that teachers measure students' vocabulary and provide advice regarding the appropriate level of books to read so that they can improve their vocabulary more systematically. Based on the effect of the extensive reading and positive feedback from the students toward extensive reading, Cha recommended integrating extensive reading into EFL reading classes.

Fujita and Noro (2009) investigated the effects of 10-minute extensive reading activities on reading speed, comprehension, and motivation with 76 first-year Japanese high school students who read graded readers extensively for ten minutes at the beginning of each class once a week over ten weeks. The participants read an average of 2,517 words (range = 336 to 11,129 words). Two versions of the EPER (Edinburgh Project on Extensive Reading) test were administered as pre-and posttests to measure reading speed and comprehension. The results of paired samples *t*-tests indicated that there were statistically significant differences on both reading comprehension and speed ( $t = 2.40, p < .05$ ;  $t = 3.31, p < .01$ ). However, the authors concluded that, based on less than one point improvement on the reading comprehension posttest, the participants made little improvement in reading comprehension.

Cirocki (2009) conducted one of the few extensive reading studies in Europe (in Poland) in his investigation of the effect of extensive reading on EFL learners' communicative competence. The participants were four male and 16 female secondary school students who attended three English classes per week over six weeks. They spent two thirds of the class time on extensive reading. During the rest of the class time, a course textbook was used. Self-made reading speed tests were

administered as pre- and posttests. The results of the reading speed tests indicated that all the students improved on the posttests; however, the author only looked at mean scores and did not conduct statistical analyses, so it was not clear how the author examined the learners' communicative competence. Despite this problem, he concluded that extensive reading effectively improves learners' reading speed.

Lin (2010) explored possible differences between 78 high school students' (38 male and 40 female students) comprehension of graded readers and the amount of time and the number of strategies they utilized to comprehend the texts over ten weeks. The same reading comprehension test was administered as a pre-and posttest. Each student read three graded readers, starting from level 2 (600 headwords) and then moving to level 3 (1,200 headwords) and level 4 (1,700 headwords). Independent samples *t*-tests were conducted to investigate reading comprehension gains and possible gender differences. The independent variable was gender. The results indicated that the female students' scores increased significantly ( $t = -2.35, p = .02$ ), suggesting that female students' reading comprehension increased more than the male students', a finding that is supported by previous literature (Logan & Johnston, 2009; Lynch, 2002). After analyzing the results of a post-reading questionnaire asking about specific strategies students used to comprehend texts, Lin

concluded that female students used their lexical, syntactic (she called them word and sentence knowledge), and world knowledge better than the male students to comprehend the texts. Lin also suggested that instructors should move away from a word-by-word translation approach and instead introduce students to extensive reading, as it leads students to comprehend stories and utilize their general schemata.

In sum, three issues regarding extensive reading in high school contexts emerged in the above studies. First, the high school students were able to read more books when they engaged in extensive reading compared with traditional approaches to teaching reading. For example, the high school students in Takase (2007a) read an average of 71,653 running words. Second, due to constraints concerning what can be taught in regular classes, some extensive reading treatments took place during summer or winter vacations because the teachers had the freedom to have students engage in extensive reading at those times. Unfortunately, this made the treatment period relatively short. Finally, half of the studies conducted in a high school context did not include a control group. This lack is a serious design flaw that makes it impossible to clearly know the effects of the extensive reading treatment.

## University Context

Two studies conducted in Japan with university students (Mason & Krashen, 1997; Robb & Susser, 1989) reported better results for extensive reading than traditional instruction involving reading a text, answering comprehension questions, and doing translation exercises. Mason and Krashen (1997) conducted three experiments. In the first experiment, they investigated whether students who had failed their English classes the previous year could improve their reading skills by engaging in extensive reading for one semester. Two intact classes took part in the study. The first, a class of “retakers” ( $n = 20$ ), who had failed their previous English course, read extensively. These students read an average of about 30 books. The control group ( $n = 20$ ) received traditional instruction including intensive reading selections, comprehension questions, and translation exercises. A 100-item cloze test was administered as a pre- and post-test at the beginning and end of the semester to both groups. Test-retest reliability with a different sample was .87. The results indicated that the experimental group performed significantly better than the control group,  $t = 2.27$ ,  $df = 38$ ,  $p < .025$ . The authors concluded that extensive reading is effective even for retakers who were initially reluctant to study English. They noted

that some students' willingness to read in English was so strong that they read not only easy books but also more challenging books.

In the second experiment, the researchers investigated the reliability of the effect of extensive reading and expanded the scope of the first experiment at a four-year university and a junior college. In each school, one experimental group ( $n = 40$  and  $31$ , respectively) engaged in extensive reading, and one control group ( $n = 39$  and  $18$ , respectively) was given traditional instruction in which they engaged in intensive reading of short texts requiring a great deal of dictionary work. The same cloze test used in experiment 1 was given as a pre- and posttest at the beginning and the end of one academic year. The results showed that the experimental groups made significantly greater gains than the control groups ( $t = 4.991$ ,  $df = 72$ ,  $p < .001$ , for university;  $t = 5.035$ ,  $df = 46$ ,  $p < .001$ , for junior college).

In the third experiment, the researchers investigated the relationship between extensive reading and writing. Along with the same cloze test used in the previous two experiments, summary writing was introduced. Three groups were formed: an English response group ( $n = 36$ ), which wrote a short summary of the book in English after extensive reading, a Japanese group ( $n = 36$ ), which wrote a short summary of the book in Japanese after extensive reading, and a control group ( $n = 37$ ), which

worked on cloze exercises. After one academic year, the results of paired sample *t*-tests indicated that both of the extensive reading groups outperformed the control group on the cloze test, ratings of summaries written in English, and time taken to read a text. Overall, the authors concluded that extensive reading was superior to traditional instruction in terms of developing reading comprehension, reading fluency, and writing skills.

Robb and Susser (1989) compared the improvement of reading comprehension by participants taught using a skills-based approach or extensive reading. The authors listed six research-based studies of extensive reading and designed their study while accounting for previous shortcomings, such as not indicating how much students read and not documenting what texts were read. The participants were four groups of first-year Japanese English majors ( $N = 125$ ) who were randomly placed into two extensive reading groups ( $N = 63$ ) and two skills groups (control groups,  $N = 62$ ). The authors did not mention the exact number of participants in each group, but I was able to obtain the information by contacting one of the authors. Two forms of the Multiple Skills Series Midway Placement Test (Boning, 1977), which was designed to measure general skills, specific skills such as making inferences, and reading speed, were administered as a pre- and posttest. An

ANCOVA was run with extensive reading groups and skills groups as the independent variable, test scores as the dependent variable, and scores of 20 items of getting the facts and scores of 10 items of making inferences as the covariate. The results indicated that the extensive group significantly outperformed the skills group ( $p < .05$ ) in general skills and reading speed. The authors claimed that the difference might have appeared because the extensive reading group spent more time at home reading. They concluded that extensive reading is an effective way for students to learn to read English enjoyably whereas translation or skills classes do not provide students with interesting materials and are therefore often perceived as boring.

Zimmerman (1997) investigated the effects of skills-based instruction over 10-weeks during which the students had 24-25 hours per week of instruction. The participants were 35 ESL students, 18 of whom were in the experimental group and 17 of whom formed the control group. The instruction included reading, composition, oral language, and academic skills. The experimental group received three hours of interactive vocabulary instruction per week. The participants were told to read self-selected materials every week and write reading records. A 50-item checklist test was administered as a pre- and posttest with items taken from the University Word List (Xue & Nation, 1984). The independent variable was treatment and the dependent

variable was the checklist posttest. The ANCOVA, which controlled for self-selected reading and required reading, indicated a statistically significant difference,  $F(1, 30) = 23.995, p = .00$ ; thus, the vocabulary instruction had a significant effect on the test scores. The author concluded that vocabulary instruction combined with reading can lead to increases in students' lexical knowledge. She further noted that the experimental group read 50% more than the control group even though they all had received the same instruction. The results of this study, which combined reading with other instruction, are similar to the results of Nakanishi and Ueda (2011) in that the extensive reading with shadowing group read more than the extensive reading only group.

Hayashi (1999) conducted a study comparing extensive reading with reading strategies. The participants were 100 Japanese university students who were divided into two levels based on their initial ITP TOEFL pretest score gathered in April. The same test was administered as a posttest in December. The scores for the listening, structure and writing, and reading comprehension section were calculated separately. The author looked at the difference between the pretest and posttest scores and concluded that the reading section scores improved more than those of the other sections. What is interesting about this study is that a questionnaire was administered

asking not only how many books in English the participants read, but also how many books in Japanese they read during the period. The results indicated that the intermediate level participants read 5.3 books in English and 13.3 books in Japanese, whereas the participants in the beginning level read 4.4 book in English and 6.6 books in Japanese. Based on the results, Hayashi concluded that extensive reading allows learners to utilize prior knowledge, it is a learner-centered approach, reading many L1 and L2 books gives readers rich background knowledge, it develops their ability to guess the meaning of unknown words from context, readers have the autonomy to select books, and this leads to motivation to read more books. She concluded that reading extensively both in the L1 and L2 is more important than teaching reading strategies in EFL contexts.

Lao and Krashen (2000) assessed the impact of reading popular literature on EFL reading rate and vocabulary acquisition with 130 first-year university students, 91 of whom were in an experimental group and 39 were in a control group. The participants in the experimental group were instructed to read six books in one semester (14 weeks), while the participants in the control group were taught traditional academic skills, such as oral skills, writing, and listening. Two pre- and posttests were administered: the Accuracy Level Test (Carver, 1987a) was used as a

measure of vocabulary acquisition and the Rate Level Test (Carver, 1987b) was used to assess reading rates. Paired samples *t*-tests were conducted on the scores of the two measures. The results indicated that there were statistically significant differences between the pre- and posttests for the experimental group ( $t = 3.88, p < .00$  for vocabulary;  $t = 11.54, p < .00$  for reading rate) while the scores of the control group were not significantly different. The results supported the outcomes of previous studies and the authors concluded that reading for meaning was what created the difference in the gains between the two groups. The results, however, need to be interpreted with caution as the control group was far less competent at the outset of the experiment.

Mason (2003) investigated the effect of extensive reading on summary writing and rewriting of the corrected summaries. The participants, 104 Japanese university students, were divided into three groups: group 1 wrote summaries in Japanese ( $n = 34$ ), group 2 wrote summaries in English ( $n = 34$ ), and group 3 wrote summaries in English, received feedback, and rewrote the summaries ( $n = 36$ ). They were provided with 5,000 graded readers to read. The participants read an average of 2,300 pages, which was approximately 500,000 words over three semesters. Group 3 received feedback 25 times during the period. A 100-item cloze test and the reading

section of a TOEIC test were used to measure reading comprehension, while a writing summary was administered as a writing pre- and posttest. The gain scores for the cloze test and the TOEIC reading section were analyzed using a two-way repeated-measures ANOVA. Statistically significant differences within the three groups were found for the cloze test,  $F(1, 91) = 359.27, p = .00$  and the TOEIC reading section,  $F(1, 85) = 53.71, p = .00$ . The writing summary was assessed through an Error Free Clause (EFC) Analysis, which involved calculating the total number of grammatically correct clauses. A repeated-measures two-way ANOVA showed statistically significant differences between the pretest and the posttest,  $F(2, 1) = 38.33, p = .00$ , but not among the groups. Overall, the results indicated that no difference was found among the three groups on any of the measurements. Interviews with the participants revealed the participants in groups 2 and 3 spent twice as much time as the group 1 on reading and writing summaries; furthermore, the participants did not enjoy writing English summaries as they were difficult to write compared with Japanese summaries. The author concluded that the approaches utilized in the study did not help the participants as much as anticipated and that extensive reading combined with writing summaries in Japanese was the most beneficial of the three approaches.

Uniquely examining the effect of extensive reading in Spanish, Rodrigo, Krashen, and Gribbons (2004) conducted a study over one semester involving 33 native speakers of English studying Spanish to investigate the effect of extensive reading on increasing vocabulary, grammar, and reading comprehension. The participants were divided into three groups: the extensive reading group did assigned and self-selected reading, the reading discussion group did only assigned readings and discussed the required readings, and the control group participated in a traditional intensive reading course. All of the participants took the same three pre- and posttests. The results indicated that the gains made by the two reading groups were significantly greater than those made by the control group on vocabulary (extensive reading vs. control,  $t = 2.51, p < .01$ ) and reading (discussion vs. control,  $t = 2.08, p < .05$ ). The difference between the extensive reading and control groups for grammar were not significant ( $t = 1.79, p = .051$ ), while the one between the reading discussion and control groups was ( $t = 3.23, p < .01$ ). Finally, there were no significant differences between the two reading groups and the control group (extensive reading vs. control,  $t = 0.15, ns$ ; reading discussion vs. control,  $t = 1.35, p = .105$ ) on the cloze test. Although the authors concluded that comprehensible input-based approaches are more effective than the traditional intensive reading approach,

the results should not be interpreted at face value because all three pretest scores for the control group were higher than those for the two reading groups. For example on the vocabulary pretest, the means of the extensive reading group, the reading discussion group, and the control groups were 25.33, 20.91, and 30.87, respectively.

Similar to Mason and Krashen's (1997) study in which a retakers' class was compared with a regular English class, Mason (2007) compared the effect of extensive reading between 16 English majors and non-24 English majors. They all took a regular English class where they listened to one of Grimm's fairy tales or another story in each class. In addition, the English majors had six additional hours of English instruction per week. The participants engaged in extensive reading outside the class and were told to read 60 to 100 pages of graded readers each week, approximately two to four books a week. To assess their reading comprehension, a 100-item cloze test was administered as a pre- and posttest. An independent samples *t*-test was conducted on the gain scores of the two groups and the results indicated that there was no statistically significant difference ( $t = 1.38, p = .18$ ) despite the fact that the English majors received seven hours of English instruction while the non-English majors received one hour English instruction per week. The author concluded that "comparing one hour per week of comprehension-based instruction to several

hours of similar instruction could, however, produce the same results. In other words, it may be the case that after a certain point, instruction of any kind is less efficient” (p. 633).

Lee (2007) presented three studies investigating the effects of extensive reading on reading comprehension and vocabulary acquisition for Taiwanese university students. Each subsequent study was modified and improved based on the limitations of the previous study. The first study was designed to examine the impact of extensive reading under limited conditions in which extensive reading instruction was provided only for 12 weeks and a limited amount of reading materials (215 graded readers) were available. The 65 participants in the experimental group read extensively for the 12 weeks, while a comparison group 1 ( $n = 38$ ) used a reading textbook, answered reading comprehension questions, and completed writing exercises. The participants in comparison group 2 ( $n = 38$ ) were encouraged to read outside of class, but no reading record was kept. The participants completed the Vocabulary Levels Test (Nation, 1990) and a cloze test (Mason & Krashen, 1997) as pre- and posttests at the beginning and end of the semester. The experimental group displayed greater gains than the two comparison groups on the 2,000 word level test and the cloze test; however, the pretest scores of the experimental group were much

lower than the two comparison groups. This led the researchers to change the second study to address this limitation.

In the second study, the researchers examined the mixed impact of the self-selected reading of graded readers and assigned reading materials over one academic year. There was one experimental group ( $n = 67$ ) and three comparison groups (comparison group 1,  $n = 40$ ; comparison group 2,  $n = 45$ ; comparison group 3,  $n = 54$ ). All groups demonstrated an equivalent level of English proficiency on the pretest. During the first semester, the experimental group freely selected graded readers to read and, during the second semester, they were required to read five assigned books: *Stuart Little*, *Charlotte's Web*, *The Trumpet of the Swam*, *The Little Prince*, and *Tuesdays with Morrie*. The results of independent  $t$ -tests indicated that the experimental group outperformed the comparison groups in vocabulary growth ( $t = 4.35, p < .01$ ), but not on the cloze test. Conversations with several students revealed that some students lost interest in the second semester because they were unable to choose books freely. This issue was addressed in the third study.

The major focus of the third study was the effect of assigned reading on the same vocabulary test and the cloze test. Two experimental groups (experimental group 1  $n = 67$ ; assigned reading and experimental group 2,  $n = 41$ ; self-selected

reading), and one comparison group ( $n = 139$ ) participated in the study. The control group was the same group described in the second study. The results of a MANOVA showed that the self-selected reading group made more gains on the cloze test than the assigned reading group and comparison group. Lee concluded that this result confirms the importance of providing students with autonomy regarding the selection of reading materials, a practice that gives them the responsibility for their own learning and enhances their motivation to read.

Controlling the other possible effects from other instruction is one of the most demanding parts of research. Yamashita (2008) explored the effects of extensive reading on different aspects of L2 ability with 31 Japanese university students who participated in a 15-week extensive reading course. Like Horst, she argued that many previous extensive reading researchers did not discuss possible effects from other types of instruction or voluntary study outside of the extensive reading programs and also stated that learning that occurs outside of extensive reading programs must be controlled if researchers are to identify the genuine effect of extensive reading. Thus, in her study, the only input the participants received was from about 500 English graded readers. Some students read their own books, but the amount of reading per week (20-30 pages) was equal to one graded reader. The participants read from eight

to 19 books with a mean of 11 books ( $SD = 2.5$ ). Along with graded readers, two participants also read paperbacks. Extensive reading tests (ERT) and placement/progress tests (PPT) (a cloze test) were administered as pre- and posttests. The makers of the ERT claim that it is the only test available to assess extensive reading ability (EPER, no date). The test is made up of relatively long reading texts because the purpose is to assess extensive reading ability, and the passages are followed by yes/no, multiple-choice, short-answer, gap-filling, and gap-summary questions. The class was divided into two groups. Each group completed the two versions of the ERT and PPT. In order to investigate the differences between the pretest and posttest, repeated measures *t*-tests were utilized. A statistically significant difference was found for the ERT, but not the PPT scores. The results indicated that extensive reading affects the development of general reading ability, while linguistic subskills, such as lexical development and improved spelling, might appear in the future.

Kweon and Kim (2008) investigated the incidental acquisition and retention of vocabulary through extensive reading. Twelve Korean university students read three authentic books that contained a total of 134,013 running words and 638 pages, and read an average of 4-6 hours a day for five weeks. The authors chose authentic

books over simplified graded readers because, in terms of incidental vocabulary learning, “using authentic texts of substantial length that may contain enough repetition of words may provide more relevant results” (p. 195). The students were required to read each book carefully because they had to take a comprehension quiz at the beginning of the following class. The participants completed a vocabulary test three times; pretest, posttest 1, and posttest 2, which was given four weeks after posttest 1. The authors looked at the test results based on word class; noun, verb, and adjective. The results of a repeated-measures one-way ANOVA showed a statistically significant difference between the pretest and posttest 1 for all three word classes at  $p < .001$ . Furthermore, the words learned were retained one month later on posttest 2.

This is another representative study of the comparison between a retakers class and a regular English class. Takase (2009) investigated the effect of sustained silent reading (SSR) on unmotivated retaker students and regular English class students. The participants were 138 Japanese university students who were divided into five groups: group 1 (18 retakers), groups 2 and 3 (freshman classes, 36 students each), groups 4 and 5 (sophomore classes, 24 students each). Group 1 engaged in extensive reading over four months while other groups engaged in extensive reading over nine months. Group 1 spent 80 minutes engaged in SSR in class and groups 2

and 3 spent 30 to 40 minutes engaged in SSR in class while groups 4 and 5 had a set text to read; thus extensive reading was done outside of class for these groups. The participants were told to start with very easy books and once they were comfortable with them, to move to more difficult ones, such as leveled readers and graded readers. The participants took the Edinburgh project of extensive reading (EPER) test at the beginning and the end of the course. Paired-samples *t*-tests were conducted on the pre- and the post-EPER scores within each group. Statistically significant differences were found for all groups except for group 4. Even though group 1 was a retakers class, they enjoyed reading and read extensively. Groups 2 and 3 also read more than groups 4 and 5, indicating that the time spent on reading in class contributed to the differences. The author concluded that SSR can be one of the most important elements in an extensive reading program and it should be implemented with beginning foreign language learners.

The only research conducted in South America was reported by de Morgado (2009), who investigated the effect of extensive reading on reading comprehension and students' perceptions of extensive reading in Venezuela. Thirty university students received extensive reading instruction for 45 minutes a week along with other regular English classes for one semester while 30 students in a control group

were subject to the same English instruction only without the extensive reading instruction. Four versions of the TOEFL were administered at the beginning and end of the semester. A questionnaire was used to collect data about the students' perceptions of extensive reading. A comparison of posttest scores between the control group and extensive reading group was not statistically significant,  $t = 1.27$ ,  $df = 59$ ,  $p = .21$ . However, the results of an independent samples  $t$ -tests for each group showed that the extensive reading group made significantly greater gains on the posttest than the control group ( $t = 3.32$ ,  $df = 29$ ,  $p = .002$ , for the extensive reading group;  $t = 0.14$ ,  $df = 29$ ,  $p = .89$ , for the control group). The author concluded that extensive reading had a positive effect on the students' reading comprehension and further stated that extensive reading created greater motivation and extensive reading is worth implementing for motivational gains alone given that motivation is one of the key factors underlying successful learning.

Al-Homoud and Schmitt (2009) conducted an extensive reading study in Saudi Arabia to compare the effectiveness of extensive reading and intensive reading over a 10-week course. Seventy college students were divided into two groups (extensive reading group,  $n = 47$  and intensive reading group,  $n = 23$ ). The extensive reading group had access to 150 graded readers divided into six levels of difficulty. A

teacher brought the books into the class and the students were able to choose the books autonomously. The authors stated that it was difficult to determine exactly how much the participants read, but they estimated that the weaker readers read approximately 31,500 running words and the better readers might have read closer to 162,000 running words over the first three weeks of the program; therefore, assuming that the participants continuously read over the rest of the treatment period, the weaker readers might have read a little over 100,000 running words and the better readers about 500,000 running words. The intensive reading group read texts in *Reading Power* (Mikulecky & Jeffries, 2005, 3rd edition), answered reading comprehension questions, and practiced reading skills, such as scanning, skimming, and making inferences. All participants were measured in terms of vocabulary, reading comprehension, and reading speed using the Vocabulary Levels Test, three reading passages of the Cambridge Preliminary English Test, and the reading section of a Pre-TOEFL for reading comprehension, and three reading passages from one graded reader, respectively. The results showed that the only statistically significant difference for reading comprehension was found for the extensive reading group on the TOEFL ( $p = .037$ ) and significant results were found for both groups for reading speed. Statistically significant differences were also found for the 2,000-word level,

3,000-word level, and academic vocabulary on the Vocabulary Levels Test between the pretest and posttest within the same group, but there was no difference between the groups. Acknowledging the limitations of the study, such as the short period of treatment and small sample size, the authors concluded that the extensive reading approach was at least as good as the traditional intensive reading approach. They also argued that extensive reading is better than intensive reading as it can create a relatively stress-free environment for students.

Yamamoto (2011) examined Japanese university students' productive vocabulary growth through extensive reading. Three tests, the Vocabulary Levels Test (Nation, 1990) for receptive vocabulary, the Productive Vocabulary Levels Test (Laufer & Nation, 1999) for controlled productive vocabulary, and the English Web VocabProfile (Cobb, 2010) for free-productive vocabulary, were administered twice at the beginning and end of a semester. All of the students ( $N = 67$ ) spent class time on intensive reading and explicit vocabulary instruction, but 33 students were in an extensive reading group. They were instructed to read at least five books outside of class over the 13-week treatment. Paired samples  $t$ -tests were conducted for both groups on the three tests. The results for the Vocabulary Levels Test ( $t = .75, p = .46$  for ER group;  $t = 1.00, p = .32$  for control group), and all the levels of the English

Web VocabProfile except for the K2 level ( $t = 2.89, p < .0125$  for the extensive reading group;  $t = 1.98, p > .0125$  for control group) were not statistically significant.

Total scores of the Productive Vocabulary Levels Test for the control group were statistically significant ( $t = -1.72, p > .0034$  for the extensive reading group;  $t = -3.62, p = .001$  for control group). (Note. The  $p$  value was set at .0034 after a Bonferroni adjustment was made.) The author concluded that the extensive reading group was able to retain lexical knowledge better than the control group. She also argued that extensive reading is generally utilized to increase an amount of vocabulary, but the results indicated that it might be used to help learners retain vocabulary, and in the long run, it can help students make connections among words they have already acquired.

Nakanishi and Ueda (2011) examined the effect of extensive reading on reading comprehension with 89 first-year Japanese university students who were divided into four groups. Two experimental groups consisted of an extensive reading only group ( $n = 20$ ) and an extensive reading and shadowing group ( $n = 22$ ). The students attended 30 classes of extensive reading class over one year. The only difference among the two was that the extensive reading and shadowing group practiced shadowing approximately 20 to 30 minutes at the beginning of each class.

All of the participants read inside and outside the class. The students in the two control groups ( $n = 21$  and  $n = 24$ , respectively) attended traditional translation-based classes. The Secondary Level English Proficiency Test (SLEP) was administered to measure students reading comprehension. Different forms of the test were administered at the beginning of a first semester, at the end of the first semester, and at the end of a second semester. A two-way between-subjects ANOVA was conducted to evaluate the effect of the treatment. The results indicated that the three tests were statistically different from each other,  $F(2, 84) = 50.75, p = .00$ , but group differences were not found,  $F(6, 168) = .17, p = .98$ . Although no group difference was detected, the authors concluded that the extensive reading groups made steady improvement, as shown by the scores on the three tests. Based on the fact that the extensive reading and shadowing group read more words and books than the extensive reading only group, it could be speculated that introducing shadowing into extensive reading might further enhance the effect of extensive reading.

In a recent study, Beglar, Hunt, and Kite (2012) investigated the effect of pleasure reading on reading rate with 97 first-year Japanese university students divided into four groups: one intensive reading group ( $n = 17$ , control group), and three pleasure reading groups. The control group engaged in usual traditional

intensive reading activities such as translating passages from English into Japanese and answering reading comprehension questions. The pleasure reading groups consisted of group 1 ( $n = 23$ ), group 2 ( $n = 22$ ), and group 3 ( $n = 35$ ). The participants were provided with six levels of graded readers and three unsimplified paperbacks over two semesters (one academic year). They all chose to read whatever they wished from books provided. The reading rate and comprehension of reading rate passages were measured through a 32-item reading rate test at the beginning of the first semester and at the end of second semester. The authors conducted a series of a hierarchical multiple regression analysis and compared effect sizes from gain scores. The results indicated that reading simplified books help students' reading rate and provided motivating effect for students as previous literature have illustrated. Although the results showed positive effects for pleasure reading, the authors argued that the results should be interpreted with some caution. First, it was only a one-year treatment, so alternative outcomes might occur if the study were conducted over a longer period of time. Second, L1 differences should be considered as Japanese is linguistically distant from English; therefore, learners with L1s genetically related to English might progress at a different rate than the Japanese participants in this study.

Many of the studies described dealt with reading speed and reading comprehension; however, one of the major differences between high school students and college students is the motivation of the students. While in high school, many Asian students must study to pass university entrance examinations; thus, they are motivated to read and to improve their reading ability. In addition, the literature reviewed here revealed two key factors concerning extensive reading. First, foreign language learners value autonomy. If books are assigned by instructors, the students tend to read less. For this reason, providing a large collection of books for students to choose from is crucial to the success of extensive reading instruction. Second, many studies indicated there is potential for stronger effects if extensive reading is combined with other forms of instruction, such as vocabulary or shadowing instruction.

The number of studies conducted in university contexts is far greater than those conducted in other contexts in part because university teachers generally enjoy more flexibility where course design is concerned and they are more likely to have training as a researcher. Studies involving university students were conducted in locations other than Asia, such as Saudi Arabia and Venezuela, a fact that demonstrates the diversity of research that has been conducted in university contexts.

## **Studies with Adult Participants**

Bell (2001) investigated the effect of extensive reading on reading speed and reading comprehension in Yemen by dividing 26 young adult students into two groups: an extensive reading group ( $n = 14$ ) and an intensive reading group ( $n = 12$ ). The experimental group read class readers and graded readers over two semesters while the intensive reading group read short texts followed by reading comprehension questions, multiple-choice questions, and gap-fills. Independent samples  $t$ -tests were conducted on the posttests for reading speed and reading comprehension. The results indicated that the extensive reading group outperformed the intensive reading group on both measures ( $t = 5.70, p < .001$  for reading speed;  $t = 7.40, p < .001$  for reading comprehension). The author concluded that extensive reading is more effective than intensive reading and that ER appears “to liberate the learner from slow reading speeds” (p. 6). Two limitations of the study were the small sample size and no information about the reliability of the instrument.

Yang (2001) investigated the performance of two groups of students in Hong Kong who read mystery novels in addition to their normal coursework. Four adult ESL classes ( $N = 120$ ) were chosen for the study. Two classes read the normal textbook, read two Agatha Christie novels, and participated in class discussions. They

read approximately 40 pages a week. The two books included 72,321 and 50,426 running words, respectively; therefore, the participants should have read 122,747 running words in total. The other two classes only read the textbook for 15 weeks. The participants completed a 100-item multiple-choice test as a pre- and posttest assessing their knowledge of English grammar, sentence structure, and usage. The mean posttest scores of the two reading groups were 75.3 and 73.7, and the means of the other two groups 66.7 and 67.0. A one-way ANOVA with the posttest scores as the dependent variable and groups as independent variable was performed. The results indicated that the two reading groups performed statistically better than the other two groups. Yang noted that the two reading groups had more time-on-task due to the writing assignments they completed; however, the proficiency gains indicated "...that the extra time on reading in English is time well spent" (p. 460).

Horst (2005) investigated vocabulary growth through extensive reading with 21 adult immigrant learners studying in Canada. She first identified the shortcomings of previous research of extensive reading and argued that in most extensive reading studies, the participants were taking other English classes in addition to engaging in extensive reading. She therefore claimed that it was not possible to conclude that the gains reported in previous studies were only from the effects of extensive reading. In

order to avoid the influence from other English classes, Horst scanned the graded readers her participants read, created lexical profiles, and wrote pre-and posttests designed to measure the participants' knowledge of the frequent and infrequent words in the books. Seventeen of the 21 participants took the vocabulary tests at the beginning and the end of the six-week extensive reading program. The results were as follows: frequent words (Pretest;  $M = 41.35$ ,  $SD = 5.38$ , Posttest;  $M = 47.94$ ,  $SD = 1.89$ ) and infrequent words (Pretest;  $M = 33.80$ ,  $SD = 8.18$ , Posttest;  $M = 43.59$ ,  $SD = 4.30$ ). Paired samples  $t$ -tests indicated that the participants made significant improvement on their knowledge of both frequent and infrequent words. The author claimed that the proposed methodology to investigate the effects of extensive reading was successful.

In addition to work with second language learners, Greenberg, Rodrigo, Berry, Brinck, and Joseph (2006) also investigated how an extensive reading program can be utilized in a classroom for native English speakers who have reading difficulties. The participants were 22 native English speakers and five non-native speakers whose average age was 39. The participants had an average of 71 hours of reading instruction; approximately five learners were in classes that lasted two hours a day, four days a week, for about 13 weeks (about one semester). The authors did not state

how much the participants read. It is common for extensive reading teachers to instruct students to read from ten minutes to one hour, but in this study the students read two hours a day; therefore, they should have read relatively large amount. The results of paired *t*-tests indicated that 27 participants showed gains in the areas of reading fluency and expressive vocabulary. Many of the participants enjoyed the program and experienced the joy of reading.

Rezaee and Nourzadeh (2011) investigated the effect of extensive reading on students' processing abilities using a self-made reading comprehension test, which had a KR-21 reliability estimate of .74. A total of 51 adult students (experimental group,  $n = 26$ ; control group,  $n = 25$ ) in Iran took part in the project. The experimental group students read books from the *Readers Series*. They were instructed to read one book a week over four weeks. The participants took the reading comprehension test, which consisted of 52 items including 24 items described as bottom-up processing items and 28 items they termed top-down processing items. The results of independent samples *t*-tests showed that there were statistically significant differences on both types of items between the experimental group and the control group ( $t = 2.28, p = .027$  for bottom-up processing items;  $t = 2.48, p = .016$  for top-down processing items). The results produced by the ER treatment were

impressive and the authors concluded that the results provide evidence that ER should be incorporated in L2 reading classes. Finally, the authors acknowledged that the limitations of the study were the short period of time (four weeks) of the study and small number of books read (four books).

The studies with adult participants provided further evidence of positive outcomes that stem from extensive reading. One of the reasons that these studies produced positive outcomes might come from the fact that the adult participants were particularly motivated to learn. They are often taking English courses voluntarily because they want to learn, and to the degree that this is true, they might provide an indication of the upper limits of the effects of extensive reading.

### **Chapter Summary**

Previous researchers (Day & Bamford, 1998; Horst, 2005; Iwahori, 2008) have summarized the results of numerous extensive reading studies in tables (see Appendices 1, 2, and 3). As shown in the table Appendix 3 in Iwahori's (2008, p. 71) study, extensive reading studies have produced statistically significant differences in areas such as vocabulary, writing, reading comprehension, reading speed, affect, and spelling, with the exception of Taguchi et al. (2004). Although the empirical findings

suggest that second language learners should engage in extensive reading, some of the findings require careful interpretation because of design and analysis problems. For example, Robb and Susser (1989) did not present essential statistical information, such as standard deviations and the sample sizes for each group, and Greenberg et al. (2006), Horst (2005), and Iwahori (2008) failed to include a control group in their studies. Two more recent studies on extensive reading (Kweon & Kim, 2008; Yamashita, 2008) are other representative cases of studies without control groups. The well-known Elley and Mangubai study was excluded from this meta-analysis because of attrition in sample size; the sample sizes were 81 for the shared book group, 98 for the silent reading group, and 121 for the control group when Class 4 took the pretest, but they were reduced to 71, 84, and 65, respectively on the posttest.

In addition to the three tables mentioned above, Wang (2010) also provided detailed tables of extensive reading research. Table 1 summarizes studies included in this meta-analysis that are not included in the Appendices.

As seen in the above studies, the researchers simply listed empirical studies of extensive reading, enumerated significant findings, and commented on the positive results of the extensive reading programs. However, it is difficult to know exactly

what contributed to the overall effectiveness of extensive reading and what moderating variables might have influenced the results. Thus, each finding on the same topic must be synthesized to reach conclusions that are more reliable than what any single study can provide.

Table 1. *Summary of Extensive Reading Studies*

Study	Population	N	Areas of interest	Results	Design
Al-Homoud & Schmitt (2009)	EFL, university, Saudi Arabia	70	RS, RC, V	Gain	E & C
Beglar et al. (2012)	EFL, university, Japan	97	RS, RC	Gain	E & C
Cha (2009)	EFL, high school, Korea	20	RS, RC, V	Gain	E & C
de Morgado (2009)	EFL, university, Venezuela	60	RC	Gain	E & C
Fujimori (2006)	EFL, high school, Japan	114	RC, O	Gain	E & C
Fujita & Noro (2009)	EFL, high school, Japan	68	RS, RC	Gain	E
Greenberg et al. (2006)	Native, adults, USA	27	RS, RC, V	Gain	E
Hayashi (1999)	EFL, university, Japan	40	RC, V	Gain	E
Iwahori (2008)	EFL, high school, Japan	33	RS, RC	Gain	E
Kweon & Kim (2008)	EFL, university, Korea	12	V	Gain	E
Lao & Krashen (2000)	EFL, university, Hong Kong	130	RS, V	Gain	E & C
Lee (2007) Study 1	EFL, university, Taiwan	103	V	Gain, but not different from control groups	E & C
Lee (2007) Study 2	EFL, university, Taiwan	206	V	Gain	E & C
Lee (2007) Study 3	EFL, university, Taiwan	180	V	Gain	E & C
Lin (2010)	EFL, high school, Taiwan	78	RC	Gain	E
Mason (2003)	EFL, university, Japan	104	RC, O	Gain	E
Mason (2007)	EFL, university, Japan	40	RC, O	Gain	E
Matsui & Noro (2010)	EFL, junior high Japan	122	RS, RC	Gain	E & C
Nakanishi & Ueda (2011)	EFL, university, Japan	89	RC	Gain, but not different from control groups	E & C
Rezaee & Nourzadeh (2011)	EFL, adults, Iran	51	RC	Gain	E & C
Rodrigo et al. (2004)	EFL, university, USA	27	V, G, O	Gain	E & C
Smith (2006)	EFL, high school, Taiwan	102	RC	Gain	E & C
Takase (2007)	EFL, high school, Japan	216	RC	Gain	E
Takase (2009)	EFL, university, Japan	138	RC	Gain	E
Tanaka & Stapleton (2007)	EFL, high school, Japan	190	RS, RC	Gain	E & C
Yamamoto (2011)	EFL, university, Japan	67	V	Gain	E & C
Yamashita (2008)	EFL, university, Japan	31	RC, O	Gain	E
Zimmerman (1997)	ESL, university, USA	35	V	Gain	E & C

*Note.* Areas of interest: RS = reading speed; RC = reading comprehension; V = vocabulary; G = grammar; O = other. Design: E & C = experimental and control group; E = experimental group only. "Gain, but not different from control groups" mean that there was no statistically significant difference in test scores between the experimental and control groups.

## CHAPTER 3

### RESEARCH SYNTHESIS

#### Meta-Analysis

The term meta-analysis was coined by Gene V. Glass (1976), who accumulated a series of studies conducted in the field of psychotherapy, analyzed the accumulated data statistically, and concluded that the treatment was generally effective. Since then, a great number of meta-analyses have been conducted. Before the emergence of meta-analysis as an approach to aggregating research findings, narrative reviews of the literature were the only way to integrate research findings. These reviews, however, could become quite subjective and meaningful quantitative findings were sometimes overlooked. Borenstein, Hedges, Higgins, and Rothstein (2009) clearly described two limitations of narrative reviews. “One limitation is the subjectivity inherent in this approach, coupled with the lack of transparency. A second limitation of narrative reviews is that they become less useful as more information becomes available” (p. xxii). Oswald and Plonsky (2010) also listed three limitations of narrative review that could be overcome through meta-analysis. The first problem is having congruous findings across studies that generally have small sample sizes, but can be considered simply because they are statistically significant. The second

problem is too much dependence on the results of traditional null hypothesis significance testing, which is a problem because the  $p$  value solely divide research findings into statistical significance or nonsignificance and it does not tell us the size or the importance of an effect. The last problem is “the limitations of the reviewers themselves” (p. 87), that is, the reviewers have too much power over what to choose and how they interpret them and those could not be the best choices for the meta-analysis.

Another traditional method for aggregating research findings is vote counting, where the number of studies with statistically significant results is counted and conclusions reached on the basis of the total number. Significance tests, however, are influenced by sample size, which means that studies with large sample sizes can more easily reach statistical significance; furthermore, significant tests do not tell the strength or magnitude of a treatment. The untrustworthiness of  $p$  values is described later in this chapter.

As described above, a quantitative approach for aggregating effect sizes across studies is more systematic and replicable than traditional, qualitative approaches because detailed information about each study is described, synthesized, and discussed. Bangert-Drowns et al. (2004) also noted that meta-analysis yields a

beneficial basis for reaching “important insight from what might otherwise be a confused and disparate literature” (p. 52).

Just as many statistical procedures have assumptions that must be met before conducting the analyses, so does meta-analysis. According to Lipsey and Wilson (2001):

1. Meta-analysis applies only to empirical research studies.
2. It applies only to studies that produce quantitative findings.
3. Meta-analysis is a technique for encoding and analyzing the statistics that summarize research findings as they are typically presented in research reports (p. 2).

In addition, the findings must not come from different research designs using different statistical approaches. Provided that the quantitative findings are comparable, different studies can be compared by transforming their results into certain forms.

Based on the above general principles, the prospective meta-analyst must formulate detailed written requirements of the criteria each study must fulfill to be included in the meta-analysis. Lipsey and Wilson (2001) listed the following seven categories for consideration: distinguishing features, research respondents, key variables, research designs, cultural and linguistic range, time frame, and publication type (pp. 16-17).

These seven categories are examined in detail in the Methods Chapter.

With specific research questions in mind, the next step in a meta-analysis is to retrieve all related studies. The most well accepted approach is to search Internet databases, such as ERIC, PsycINFO, and EBSCO, and academic databases using keywords. The outcome of the search is often enormous; therefore, special attention should be given to choice of keywords. If the keywords are too narrow, eligible studies cannot be identified. The candidate studies, then, are screened through criteria for inclusion and the references of those studies are examined to identify more candidate studies.

There are two primary ways to compare independent studies. One is to make use of  $p$  values (Becker, 1994). Individual researchers frequently evaluate the results in terms of their statistical significance (e.g.,  $p$  levels). Researchers often view statistical significance as a positive outcome and nonsignificance as an undesirable one; however, a  $p$  should not be treated at face value, as  $p$  can be employed by utilizing  $Z$ , the standard normal deviate corresponding to the  $p$  value. Rosenthal (1987) defined  $Z$  as

$$\frac{Z_1 - Z_2}{[2]^{1/2}} \text{ is distributed as } Z.$$

Then, the  $p$  value associated with the obtained  $Z$  is obtained. The  $p$  value can be checked to determine if there is a statistically significant difference between the studies. However, a problem with this approach is that statistical significance is partly a function of sample size; thus, the magnitude of the influence of an intervention or treatment cannot be used as a simple measure of the effect of the intervention (Murphy & Myers, 2004; Valentine & Cooper, 2003). The simple equation is (Rosenthal & DiMatteo, 2001, p. 63):

$$\text{Significance Test} = \text{Effect Size} \times \text{Study Size}$$

Abelson (1997) noted that if wombats could not meet a significant level at  $p$ -value of .07 and dingbats could produce a  $p$ -value of .05, we might conclude that the treatment effect was effective only for dingbats, but in fact, the treatment effect might have been similar. Furthermore, if the sample size of wombats was smaller, the treatment effect for the wombats might have been larger. McCartney and Rosenthal (2000) stated that “policymakers as well as researchers sometimes mistakenly confuse the significance of a finding with the importance of a finding. A significance test that results in a small  $p$  value does not necessarily connote an important finding” (p. 173).

Given the untrustworthiness of the use of  $p$  values, a more appropriate way to compare the statistical results of independent studies is to compute effect sizes (e.g.,  $d$ ,

$g$ ,  $\Delta$ , or  $r$ ). The American Psychological Association (APA) (2010) has urged “for the reader to appreciate the magnitude or importance of a study’s findings, it is almost always necessary to include some measure of effect size in the Results section” (p. 34). The effect size statistics standardize various forms of quantitative research findings, and this makes it possible to compare different studies. The equations for three effect sizes are (Rosenthal & DiMatteo, 2001, p. 71):

$$\text{Cohen's } d = \frac{M_1 - M_2}{\sigma_{pooled}}$$

$$\text{Hedge's } g = \frac{M_1 - M_2}{S_{pooled}}$$

$$\text{Glass's } \Delta = \frac{M_1 - M_2}{S_{control\ group}}$$

The results of meta-analytic procedures are evaluated in terms of their effect sizes.

For each of the studies to be compared, effect sizes need to be calculated. For example, from the effect size  $r$ , the associated Fisher  $z_r$  can be calculated with the following formula:

$$\frac{1}{2} \log_e [(1+r)/(1-r)]$$

Where  $\log_e$  is the natural logarithm,  $r$  is the correlation coefficient. Then, from the associated Fisher  $z_r$ ,  $Z$  can be computed using the following formula:

$$\frac{z_{r1} - z_{r2}}{[1/(N_1 - 3) + 1/(N_2 - 3)]^{1/2}} \text{ is distributed as } Z$$

Where  $z_{r1}$  is the associate Fisher  $z_r$  in Group 1,  $z_{r2}$  is the associate Fisher  $z_r$  in Group 2,  $N_1$  is the number of subjects in Group 1,  $N_2$  is the number of participants in Group 2. Finally again, the  $p$  value associated with the obtained  $Z$  is obtained. The  $p$  value can be checked to determine if there is a statistically significant difference between the studies.

The same principle applies when comparing any number of independent studies. Combining any number of independent studies to determine more accurate effects of treatments is the fundamental purpose of conducting a meta-analysis.

Rosenthal and DiMatteo (2001) stated “Meta-analysis allows researchers to arrive at conclusions that are more accurate and more credible than can be presented in any one primary study or in a nonquantitative, narrative review” (p. 61). In sum, meta-analysis shift the emphasis from single studies to multiple studies and focuses attention on the practical importance of the effect size instead of the statistical significance found in single studies. It also increases the power of the analyses by greatly increasing the sample size by adding the participants in multiple single studies to form a single large sample.

## Criticisms of Meta-Analysis

All statistical procedures have weaknesses, and meta-analysis is no exception. Two common problems are the file drawer problem and publication bias. Researchers can lose interest in a study when they do not find statistically significant results and this can lead them not to publish the study. Rosenthal (1979) called this the “file drawer” problem, meaning that studies with non-significant results frequently end up not being published. It is occasionally referred as fugitive literature. Statistically significant results increase the probability of a study being published, as such studies are treated as important, while studies with non-significant results are often treated as unimportant. Papers conventionally appear in a journal when researchers obtain significant results; however, data are wasted when papers are rejected because researchers could not obtain significant results even in well-designed studies. Hunt (1997) also described “the survey of psychological research articles just mentioned found that researchers were eight times more likely to submit their results to a journal if they were statistically significant than if they were not” (p. 119). To solve this problem, journal reviewers need to be aware of the file drawer problem and identify well-designed studies, including those with non-significant results, which are

unpublished so that readers are informed of those outcomes in a given field.

Ultimately, this approach helps reduce a second problem, publication bias.

The second problem, publication bias, refers to a situation where “the research that appears in the published literature is systematically unrepresentative of the population of complete studies” (Rothstein, Sutton, & Borenstein, 2005, p. 1). It usually comes from missing studies. To avoid publication bias, Pigott (2012) argued that “one strategy for addressing publication bias is the use of thorough search strategies that focus on published, unpublished and fugitive literatures” (p. 80). Thus, multiple databases, books, and journals must be examined to identify all existing studies. In this study, publication bias was investigated through the use of a sensitivity analysis.

File drawer problems and publication bias create a quality issue for researchers wishing to conduct a meta-analysis. If only published papers with statistically significant results are utilized in meta-analyses, the results are positively biased, as studies with non-significant results are not included. In such cases, the distribution of effect sizes is skewed or cut off. To make matters worse, Lipsey and Wilson (2001) stated “it is known that the effects reported in published studies are generally larger than those reported in unpublished ones” (p. 19). This tendency is

prevalent in various research fields; however, it might be ameliorated once researchers acknowledge the importance of meta-analysis, as it might mean that more well-designed studies, whether they produced significant results or not, will be published.

The biggest concern regarding meta-analysis is the mixture of studies that is included in the analysis; this has been referred to as the “apples and oranges” problem. The argument is that studies originate in numerous journals, some of which are higher quality than others. This fact suggests that the summary effect might ignore important differences across studies. Thus, when aggregating studies, researchers must first identify studies in which the same topic was investigated, and then include and exclude studies based on the stipulated criteria. The criteria need to be strict and well justified to cope with the above criticism.

Norris and Ortega (2006b, pp. 6-7) listed three features of a research synthesis:

1. Research synthesis always includes an explicit articulation of how the relevant literature was searched and how primary studies were selected for review.
2. Research synthesis focuses on the variables, characteristics, and data reported in primary studies, rather than on the study-specific conclusions offered by primary researchers.

3. A research synthesis compiles findings and seeks generalizations by examining categories of data and methodology that cut across studies, in order to create as systematic a depiction as possible about what we know, what we do not know, and why.

These three features are achieved in this study as I (a) collect articles based on the stipulated criteria; (b) focus on specific variables, characteristics, and data reported in primary studies; (c) examine findings from the accumulated research to depict known and unknown aspects of extensive reading; and, (d) suggest avenues for future research.

### **Meta-Analytic Studies in the Field of Second Language Acquisition**

In this section, the meta-analysis procedures employed by second language acquisition (SLA) researchers and, specifically, the effect sizes they have used are described. The first meta-analysis study conducted in the field of SLA was by Sahari (1997), who reviewed 29 studies with a total sample of 3,005 participants to investigate the effects of elaboration from text-based information. Given the fact that previous studies on elaboration were inconclusive, researchers beliefs about the effectiveness of elaboration differed. The results showed that elaboration enhances

text-based learning and that the learners created a connection between text-based information and existing knowledge.

Ross (1998) investigated self-assessment in second language testing by conducting a meta-analysis for all the four skill areas from 60 correlations reported in ten studies in the second language testing literature. The most common approach to self-assessment is to examine correlations between self-assessment scales and each specific skill. Utilizing the effect size  $g$  from standardized correlation coefficients, he first investigated the overall effect of self-assessment. The average correlation for the 60 self-assessments was  $r = .63$ ; however, the homogeneity test indicated that there is a wide variation across the samples, thus he continued to explore correlations of self-assessment with four separate skill areas. For the reading skill, with 23 effect sizes, the average correlation was  $r = .61$ . For the listening skill, with 18 effect sizes, the average correlation was  $r = .65$ . For the speaking skill, with 29 effect sizes, the average correlation was  $r = .55$ . For the writing skill, with 15 effect sizes, the average correlation was  $r = .52$ . He concluded that there is enough evidence of strong correlations between self-assessment and four skill areas.

Blok (1999) reviewed ten studies concerning with reading to children. The dependent variables were classified into two domains: oral language and reading

skills. Eleven independent samples from ten studies were collected. The overall effect sizes for oral language ( $d = .63$ ) and for reading ( $d = .41$ ) indicated that reading at school had a large effect on children's oral language development and a medium effect on reading development. Although the results seem promising, the author warned us not to take them as face value because the number of studies included was small and the quality of each study was not up to the standard.

The most well-known meta-analysis study in the field of second language acquisition is by Norris and Ortega (2000), who investigated the effectiveness of L2 instruction from experimental and quasi-experimental studies published between 1980 and 1998. Norris and Ortega identified 49 independent studies, compared the effect size  $d$  across the studies, and concluded that L2 instruction makes a difference and that explicit L2 instruction is more effective than implicit L2 instruction.

Masgoret and Gardner (2003) researched the relationship of second language achievement to five attitude/motivation variables from studies conducted by Gardner and associates and also examined two additional variables, availability of the language community and education levels. They identified 75 studies with data from 10,489 individuals, involving 56 published articles, and 19 unpublished articles or dissertations. The procedures conducted in this study were based on Hunter and

Schmidt's (2004) procedure, which requires the researcher to compute the mean correlations of each criterion with each variable using the reliability estimates for each sample, and then to aggregate everything for each variable. The results indicated that the correlations were all positive; furthermore, the correlation between motivation and self-ratings of achievement was the highest (.39) and the one between instrumental orientation and objective measures was the lowest (.08). Even though previous research showed a negative relationship between some integrative motivation measures and L2 achievement, the authors argued that the evidence from this study demonstrated otherwise.

A meta-analysis of bilingual education was conducted by Rolstad, Mahoney, and Glass (2005). Researchers investigating bilingual education in Arizona as well as nationally have concluded that bilingual education is an effective instructional approach. The authors of this study attempted to provide additional support for its effectiveness by conducting a meta-analysis focused on studies conducted in Arizona. Although only four studies were included in the analysis, they demonstrated a positive effect for bilingual education in Arizona. When the four studies were coded, a total of 43 instances of bilingual programs that compared one or more outcome measures was identified; effect sizes were calculated for all 43 instances. The results were the same

as those of previous bilingual education studies; bilingual education was superior to an English only approach. However, bilingual education was not applied in the United States for political reasons. The authors claimed that “in Arizona and elsewhere in the United States, bilingual education appears to face opposition primarily for political, rather than pedagogical, reasons” (p. 62). The authors suggested, based on the results, that the current policy of English-only education should be reconsidered.

The first collection of studies on research synthesis in SLA was published in 2006, which featured five empirical syntheses. In relation to L2 reading, reading strategy training was investigated by Taylor, Stevens, and Asher (2006). Their primary interest concerned the effect of explicit reading strategy training (ERST) on L2 reading comprehension. A number of previous studies on the use of conscious reading strategies indicated significantly better results than those without such strategies. The authors used data from 23 sample studies, and reported a mean effect of  $d = .54$ , which indicates that 68% of the students receiving ERST outperformed the students not receiving ERST. They also investigated whether other moderating variables such as age and type of strategy training had any effect to the strategy outcomes.

Other meta-analysis studies dealt with the effects of corrective feedback on L2 grammar learning (Russel & Spada, 2006), the relationship between Universal Grammar and second language acquisition (Dinsmore, 2006), the link between interaction and the acquisition of specific grammatical and lexical features (Keck et al., 2006), and the effectiveness of instruction on interlanguage pragmatic development (Jeon & Kaya, 2006). CALL empirical research on second language reading comprehension and vocabulary were meta-analyzed (Abraham, 2008), and format effects on test performance (In'nami & Koizumi, 2009). The studies reviewed above are summarized in Table 2.

Oswald and Plonsky (2010) specifically discussed meta-analysis studies in second language research including above-mentioned studies and described how meta-analysis studies were done and what it needs to conduct one. They showed that there is a growing tendency for second language researchers to conduct meta-analysis in the field of second language acquisition (see pp. 100-101 for a detailed list of meta-analyses of L2 research). Using the same line of argument, Plonsky (2012) listed references for meta-analyses that had been conducted in the field of applied linguistics. This list included about 70 meta-analyses and papers/presentations concerning meta-analysis.

Table 2. Summary of Meta-Analytic Studies in Second Language Acquisition

Author	Areas of interest	Effect sizes used
Abraham (2008)	CALL reading comprehension and vocabulary learning	<i>d</i>
Blok (1999)	Reading to children	<i>d</i>
Dinsmore (2006)	L2 learner's access to UG	<i>d</i>
Goldschneider & DeKeyser (2001)	Morpheme acquisition	Raw data were used
In'nami & Koizumi (2009)	Format effects on test performance	<i>g</i>
Jeon & Kaya (2006)	Interlanguage pragmatic development	<i>d</i>
Keck et al. (2006)	Task-based interaction and acquisition	<i>d</i>
Masgoret & Gardner (2003)	Language achievement by attitude and motivation	<i>r</i>
Norris & Ortega (2000)	L2 instruction	<i>d</i>
Norris & Ortega (2001)		<i>d</i>
Rolstad, Mahoney, & Glass (2005)	Bilingual education	<i>g</i>
Ross (1998)	Self-assessment	<i>g &amp; r</i>
Sahari (1997)	Elaboration as a text-processing	not specified
Taylor, Stevens, & Asher (2006)	Reading strategy training	<i>g &amp; d</i>

In sum, based on the meta-analytic studies conducted in the field of second language acquisition in Table 2, *d* (Cohen, 1988) is the most commonly used effect size. Rosenthal (1994) suggested that “it seems natural to employ *d*-type effect size estimates when the original studies have compared two groups so that the difference between their means and their within-group *S*'s or  $\sigma$ 's are available” (p. 236); therefore, in this study, I utilized *d* as the standardized effect size. The calculation and interpretation of effect sizes are described in the Method section.

## **Meta-Analyses of Extensive Reading**

A review of L1 extensive reading was published in the National Reading Panel (NRP) report (2000). Fourteen studies were included in the analysis based on strict criteria; however, these collected studies used different moderator variables and differed in methodological quality; thus, a meta-analysis could not be performed. Instead, the NRP claimed “examined individually” (p. 3), which means that they were narratively reviewed. The conclusion the NRP came to was that no positive effect of extensive reading was found on reading comprehension; however, the results need to be interpreted with caution (Grabe, 2009; Kuhn & Stahl, 2003).

There are three meta-analyses on L2 extensive reading. First, Krashen (2007) collected studies with adolescents and young adults and calculated 13 effect sizes for cloze tests and nine effect sizes for reading comprehension, then reported the average effect size for both measures ( $d = .70$ ). Krashen concluded that “in-school self-selected reading is effective and its effects are robust” (p. 27). However, the results need to be interpreted carefully, as detailed information, such as how the researcher searched for studies, criteria inclusion, effect size calculation, and a literature review of the studies included in the analysis were not mentioned in the paper. Furthermore,

it is the shortest meta-analysis of the three, as it is only five pages excluding references.

Wang (2010) conducted what was probably the most comprehensive meta-analysis among the three reviewed here. He focused more on moderator variables than on the overall effect of extensive reading. Thirty five studies from 1977 to 2007 were collected and divided into groups using six moderator variables: reading speed, vocabulary acquisition, reading comprehension, writing fluency, accuracy, and language proficiency. Analyses were conducted separately for controlled and uncontrolled experiments. The results of the uncontrolled experiments indicated that large effects were obtained for reading speed, reading comprehension, writing fluency, accuracy, and language proficiency. On the other hand, the results of the controlled experiments showed that a large effect for reading speed and a medium effect for reading comprehension were found. College learners benefited more than young learners where accuracy was concerned. Wang concluded that even though there are uncertainties with the uncontrolled experiments, extensive reading has a positive effect on language learning and it should be a part of a foreign language curriculum. This study is discussed in more detail with the results of this study in Chapter 6.

Kim (2011) presented a paper at the first Extensive Reading World Congress held in Japan in 2011 that appeared to be a meta-analysis of classroom ER research. A request was sent to the author for more information about the presentation, but there was no response, so a review of the study is not possible.

The current study differs from the above three meta-analyses in four ways. First, this meta-analysis includes more recent studies than previous studies, as it includes studies published up to 2012. Thus the results more accurately show the effect of extensive reading compared with previous meta-analyses.

Second, the current study includes more research conducted in Japan, where extensive reading is an increasingly common approach to increasing the amount of English input available to learners. I have included as many studies as possible, each of which was collected based on the stipulated criteria so the analyses are based on larger samples that provide greater statistical power. One of the main reasons I have not included many studies conducted outside of Japan is that they were not present in the academic databases I searched. Although I considered removing article found in small journals such as KATE from the analyses, I ultimately included those studies provided they met the criteria for inclusion. Wang (2010) also included 16 local proceedings available only in Taiwan.

Third, the analysis of moderator variables shows how many studies were conducted in Japan and the overall results of research conducted in that context. Finally, the impact of learners' ages and the length of instruction are investigated to identify any systematic disparities among these variables.

### **Purposes of the Study**

As reviewed above, many researchers have concluded that a number of positive effects flow from extensive reading; however, the magnitude of the effects and the length of time needed to achieve certain effects are unknown. Thus, the first purpose of this study is to investigate the overall effectiveness of extensive reading by aggregating the results of a large number of extensive reading studies found by searching a number of academic databases. Maximizing the size of the sample provides more statistical power, which is one advantage of meta-analyses.

The second purpose is to investigate whether learners' ages impact learning from extensive reading differently. This purpose is important because extensive reading has been implemented with learners ranging from their early teens (i.e., junior high school students) to adult learners in their 30s and 40s.

The third purpose is to investigate whether the length of time that second language learners engage in extensive reading influences outcome measures. If an effect is found, how long is an appropriate period of time? Investigating the length of time that needs to be spent on extensive reading is beneficial for instructors when creating reading curricula. Furthermore, identifying and specifying the measurements utilized is a crucial part of this study, as this information helps future researchers who wish to use the same measurements to replicate a study or compare their results with those obtained in other contexts.

The fourth purpose is to investigate whether the amount read influences test scores, and what kinds of tests are utilized to determine the effect of extensive reading. If the amount read is influential, the results can provide guidelines concerning how long the treatment should be and how much should be read if learners are to make significant gains. Furthermore, by investigating the kinds of tests utilized in each study, the results from various studies can be compared. As can be seen in Appendices 1, 2, and 3, summaries of past extensive reading studies dealt with sample size and gains made, but the measures that were utilized in each study were not investigated. Thus, identifying and specifying the measurements utilized is a crucial part of this study as this information helps future researchers who wish to use the

same measurements to replicate a study or compare their results with those obtained in other contexts.

The last purpose of this study is to identify directions for future extensive reading research. While many studies have been conducted in certain areas, such as the effect of extensive reading on vocabulary acquisition, there are likely a number of other important areas in need of further investigation. Identifying these areas will be valuable for increasing our knowledge of the effects of extensive reading.

### **Research Questions**

This study was designed to address the following research questions:

1. How effective is extensive reading instruction?
2. Do learner's age impact learning differently?
3. Does length of instruction influence test scores? If so, how long is an appropriate period of instruction?
4. What kinds of tests are utilized to determine the effects of extensive reading?
5. What deficiencies are there in extensive reading research?

## **CHAPTER 4**

### **METHODS**

In this chapter, I describe the methodology used in this study. First, the procedures used in this study are described. Next, I explain how the articles used in this study were identified, what databases were utilized, and describe the criteria for inclusion. Finally, a detailed description of each study included in the meta-analysis is presented.

#### **Procedures**

First, primary studies on extensive reading were gathered after conducting a comprehensive search of academic databases and journals. The articles were closely inspected to ensure that they met the criteria for this meta-analysis. Second, the articles that met the criteria were coded so that the effects of moderating variables could be identified. Third, effect sizes for each study were calculated. Fourth, homogeneity of variance tests ( $Q$ -test) were calculated to determine whether the distribution of weighted effect sizes for extensive reading studies was normal and due only to sampling error. Then, all the effect sizes were aggregated to determine the

overall effect of extensive reading. Finally, the moderating variables, which are discussed below, were examined. A meta-analysis is particularly suitable for investigating moderating variables that are considered crucial, but that any single study fails to explore (Norris & Ortega, 2007).

Lipsey and Wilson (2001) claimed “if the research studies of interest involve findings in both pre-post forms and group contrast forms, those different findings should be separated, each form coded with the appropriate effect size statistic, and each of the resulting sets of effect size values meta-analyzed separately” (p. 45). As this research involves two types of forms—group contrast forms and pre-post forms—two separate analyses are conducted. Norris and Ortega (2000) and Lyster and Saito (2010) did two separate analyses based on different contrast forms. Some meta-analysts (e.g., Blok, 1999; Li, 2010) have entirely excluded studies without control groups and therefore not conducted separate analyses.

### **Searching the Literature**

The following search engines were utilized to identify articles to include in the meta-analysis: Academic Search Premier, Education Research Complete, the Educational Resources Information Center (ERIC), Directory of Open Access Journals, Dissertation Abstract, Linguistics and Language Behavior Abstracts (LLBA),

MLA International Bibliography, Professional Development Collection, ProQuest, PsycARTICLES, PsycINFO, and Scopus. The Web of Science, which was used by In'nami and Koizumi (2010), was also checked. In order to find research papers, books, and magazines in Japan, GeNii, which is a Japanese database of academic papers, books, and research reports, was also examined. Finally, as suggested in Plonsky and Oswald (2012), Google Scholar was also utilized.

First, to access all reading studies, the initial search utilized *reading* as a keyword; however, the use of this general keyword resulted in more than 32,000 studies being identified. Therefore, searches utilizing *extensive reading*, *pleasure reading* and *graded readers* as keywords were conducted, and 156 articles were identified and retrieved for review.

Second, the following 26 journals were electronically and manually examined to determine if they included related studies: *American Educational Research Journal*, *Applied Linguistics*, *Canadian Modern Language Review*, *ELT Journal*, *International Journal of Educational Research*, *Journal of Educational Psychology*, *Journal of Reading*, *Journal of Research in Reading*, *Language Learning*, *Language Teaching Research*, *Language Testing*, *Modern Language Journal*, *Reading in a Foreign Language*, *Reading Matrix*, *Reading Research Quarterly*, *RELC Journal*, *Second*

*Language Research, Studies in Second Language Acquisition, System, TESL-EJ, and TESOL Quarterly.* In addition, in order to search for extensive reading studies conducted in Japan, the following journals were used: *ARELE (Annual Review of English Language Education in Japan), JACET (Japan Association of College English Teachers) Journal, JALT (Japan Association for Language Teaching) Journal, JLTA (Japan Language Testing Association) Journal, KATE (Kanto-koshinetsu Association of Teachers of English) Journal, and LET (Language Education & Technology).*

Finally, to ensure that additional studies not identified via the above two methods were also discovered, eight books on reading in a second language (Coady, 1997; Day & Bamford, 1998; Grabe, 2009; Grabe & Stoller, 2011; Han & Anderson, 2009; Koda, 2004; Nuttall, 2005; Takase, 2010) were checked manually, and references to additional studies were cross-checked.

### **Criteria for Inclusion**

Five criteria for inclusion were employed when assessing the candidate studies.

1. The article concerned extensive reading, pleasure reading, or graded readers and was published by July 2012.

2. Students engaged in extensive reading instruction for a certain period of time.
3. The article was a report of an experimental or quasi-experimental study.
4. The means and standard deviations were reported, or *t*-values or *F*-values were presented.
5. The participants were elementary school children or older.

Lipsey and Wilson (2001) listed seven criteria for consideration when conducting a meta-analysis. Criterion 1 concerns the primary focus of the studies, which in this case were extensive reading, pleasure reading, and graded readers. It also set the time frame for the studies. Criterion 2 also concerns the time frame in terms of how long the treatment lasted. Criterion 3, which concerns the research design, stipulates that qualitative approaches are unsuitable; only experimental or quasi-experimental studies are candidates for inclusion. Criterion 4 concerns key variables that produce the data used to calculate effect sizes. Necessary information or data, such as means and standard deviations, must be presented, preferably in tables or figures. At the end of an analysis, effect sizes are aggregated and compared to yield results. Criterion 5 restricts the research respondents, so that the effects of the treatment for different age groups can be investigated.

As described in detail in the National Reading Panel (NRP) report (2000), the titles and abstracts of studies collected were examined. Finally, studies meeting all five criteria above were included in this study. Unpublished studies were not included except for dissertations. There are several arguments regarding the inclusion of unpublished studies in a meta-analysis. Authorities who are against their inclusion argue that the quality of unpublished studies is uncertain; however, this argument does not necessarily apply to doctoral dissertations, as most researchers would agree that dissertations go through sufficient scrutiny by dissertation committee members to ensure reasonable levels of quality. Indeed, some researchers include unpublished studies, including dissertations (Oswald & Plonsky, 2010). Li (2010) included 11 dissertations and 22 published studies to investigate the effects of corrective feedback on L2 morph-syntactic development, and he found no statistically significant difference in effect sizes between the two types of publications. In fact, the mean effect size for dissertations was larger than that in the published studies. This approach provided good counter-evidence against the critique that the quality of unpublished dissertations and published studies might differ considerably.

Whether it is published or unpublished, including a large number of studies in a meta-analysis is valuable because it increases the sample size and statistical power.

On the other hand, other meta-analysts avoid sampling unpublished studies because those studies can be biased and not meet the standards of scientific quality (Keck et al., 2006).

After conducting a thorough literature search, 34 studies were included after applying the above criteria; 18 of the studies included control groups and the remaining 16 studies did not. Studies preceded by an asterisk in the reference section are included in the meta-analysis.

Some researchers' studies were not included in this meta-analysis because they did not provide the statistical information necessary to conduct a meta-analysis. For instance, Robb and Susser (1989) did not report the number of students in each group and the standard deviations for each test score. Furthermore, Sheu (2003) failed to include standard deviations for the pretests and posttests, so the study was excluded from this meta-analysis. Requests were sent to the authors of these studies, one of whom provided the necessary data.

Rosenthal (1995) described the above kinds of studies when he stated that many researchers simply wrote that "there was no effect of X on Y" or "the effect was not significant." Meta-analysis is a summary of data and statistical findings, not a summary of an author's conclusions, so the statements such as those above are of

little help to the meta-analyst. However, if the meta-analyst can obtain the relevant means and standard deviations, effect sizes can be calculated (p. 184). Lee and Huang (2008) also stated that in each study the descriptive statistics, methodology, and procedures must be reported so that meta-analysts can gain a more comprehensive understanding of that study.

A word of caution should be mentioned. When collecting and reviewing studies to include in a meta-analysis, meta-analysts should be aware of studies sponsored by agencies or publishers because the results of such studies can involve a conflict of interest. None of studies included in this meta-analysis were sponsored by an agency or publisher.

## **Analyses**

### **Calculation and Interpretation of Effect Sizes**

A measure of effect size, Cohen's  $d$ , was calculated for all outcomes using the formulas below to investigate the effects of extensive reading instruction. As mentioned above, descriptive statistics are required to calculate effect sizes, but some studies did not include standard deviations or pretest scores and many authors failed to include control groups, which are needed to compare effects. Due to these

circumstances, two types of data were extracted. For this reason, two formulas for calculating effect sizes were required.

For group contrasts where experimental group and control groups were included, the following equation was used:

$$\text{Cohen's } d = \frac{M_1 - M_2}{\sigma_{pooled}}$$

where  $M_1$  is the mean of posttest of experimental group,  $M_2$  is the mean of posttest of control group,  $\sigma_{pooled}$  is average population standard deviation.

For pre-post contrasts where one group was measured at Time 1 and Time 2 the following equation (Lipsey & Wilson, 2001, p. 44) was used:

$$ES = \frac{X_{T2} - X_{T1}}{S_p}$$

where  $X_{T2}$  is the mean at Time 2,  $X_{T1}$  is the mean at Time 1,  $S_p$  is the pooled standard deviation of the Time 1 and Time 2 scores. As for the pooled standard deviations for the two contrasts, the average of the two standard deviations was used.

One alternative for calculating the above effect sizes if necessary data are missing is to utilize  $t$  or  $F$  values (see Rosenthal, 1994, for more detailed procedures for the alternative conversion).

Effect sizes can be interpreted in terms of standard deviation units. Ellis (2010) explained how to interpret effect sizes in a straightforward manner: "A score

of .50 means that the difference between the two groups is equivalent to one-half of a standard deviation, while a score of 1.0 means the difference is equal to one standard deviation. The bigger the score, the bigger the effect” (pp. 10-11). He noted that one advantage of reporting standardized effect sizes is that they can be compared across studies.

Cohen (1988) suggested the following guidelines for designating effects as small, medium, and large:

$d = .20$  or  $r = .10$  is considered a small effect size,

$d = .50$  or  $r = .30$  is a medium effect size, and

$d = .80$  or  $r = .50$  is a large effect size.

“The larger this value, the greater the extent to which the phenomenon under study is manifested” (Cohen, 1988, p. 10). Researchers need to be aware, however, that the above criteria are not definitive guidelines that apply to all the research fields equally.

Oswald and Plonsky (2010) acknowledged this issue when they stated that “Cohen’s (1988) benchmarks appear to underestimate the experimental effects generally obtained in L2 research” (p. 99). They then offered a preliminary set of SLA standards for effect sizes:  $d = .40$  represents a small effect,  $d = .70$  is interpreted as a medium effect, and  $d = 1.00$  is a large effect (p. 99). Plonsky (2011) utilized new

standards for interpreting effect sizes as well as Cohen's standards to provide a more detailed description of the potential implications of obtained effect sizes. Plonsky (2012) and Shintani, Li, and Ellis (2013) also used these new criteria. It might be time to rethink the traditional guidelines and consider using the more recently proposed guidelines in the field of SLA. Both criteria are utilized in this study.

### **Fixed-Effect Model and Random-Effect Model**

When conducting a meta-analysis, two statistical models can be used, the fixed-effect model and random-effect model (Cooper, 2010; Cornell & Mulrow, 1999; Hedges, 1994; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001; Raudenbush, 1994). In the fixed-effect model, it is assumed that there is only one effect size across all the studies in the analysis. If any difference is observed, it is considered to be due to sampling error. In the random-effect model, it is assumed that the effects differ from study to study. For instance, the participants in each study might differ in age, native language, where they live, or testing format. In addition, researchers (Ellis, 2010; Pigott, 2012) have noted conceptual advantages of the random-effect model over the fixed-effect model. Ellis (2010) argued that "given that most reviewers will be interested in making unconditional inferences that apply to studies that were not included in the meta-analysis or that have not yet been done, then the random-effects

model is unquestionably the better choice” (p. 129). Thus, the random-effect model was utilized in this study; however, as previous researchers (e.g., Li, 2010; Patall, Cooper, & Robinson, 2008) reported both models in order to offer a more comprehensive picture of the studies under investigation, the results of both models are reported.

### **Homogeneity of Variance Tests**

Homogeneity of variance tests ( $Q$ -test) were calculated to determine whether the distribution of weighted effect sizes for extensive reading studies was normal and due only to sampling error. The formula for calculating  $Q$  is as follows:

$$Q = \sum_{i=1}^k W_i (Y_i - M)^2$$

Where  $k$  is the number of studies, in which the  $i$ th study reports one observed effect size  $Y_i$ ,  $W_i$  is the study weight and  $M$  is the summary effect (See Shadish and Haddock (1994) for a more detailed description of these procedures.)

If the obtained value of homogeneity  $Q$  rejects the critical level of significance of .05, the effect sizes are considered heterogeneous. In such cases, it is necessary to investigate various features of the studies under consideration. There is a widespread belief that the researcher must stop the meta-analysis if the homogeneity

test is rejected, as it indicates that the difference was created by sampling error, not by the treatment itself; however, Cooper (2010) claimed that it is possible to proceed with the analysis “if there are good theoretical or practical reasons for choosing moderators” (p. 185). Borenstein, Hedges, Higgins, and Rothstein (2009) further suggested that “a nonsignificant  $p$ -value should not be taken as evidence that the effect sizes are consistent, since the lack of significance may be due to low power” (p. 113). Rosenthal (1995) also stated that “Contrasts, particularly planned contrasts, can and should be computed among the obtained effect sizes whether the overall test of heterogeneity is significant or not” (p. 188). Thus in this analysis, even if the homogeneity test is rejected, the analysis will be continued.

### **Descriptive Statistics**

Tables 3 and 4 summarize the descriptive statistics and effect sizes calculated for the studies included in this meta-analysis. The 36 studies provided 70 unique effect sizes (38 effect sizes for group contrasts and 32 effect sizes for pre-post contrasts, respectively). Fujimori (2006) (See Table 4) used a control group; however, she used second-year students in the control group who had had an extra year of English study compared to the first-year students in the study; hence, the study was

not properly controlled. For this reason, the study was included in the pre-post contrasts.

Some effect sizes (e.g., two effect sizes for Rezaee and Nourzadeh [2011] and Horst [2005]) were based on the same samples; thus, they are considered non-independent observations (Lipsey & Wilson, 2001; Norris & Ortega, 2006). Multiple effect sizes from the same sample contribute more than their fair share to the meta-analysis and thereby distort the results. Thus, an adjustment was applied. If a study produced two or more effect sizes from the same sample within a study, the effect sizes were averaged; however, if a study provided two or more effect sizes using the same control group but different treatment groups, the effect size from the larger sample size was chosen.

Each primary study contributed only one effect size in the calculation of the overall effect of extensive reading; however, in the moderator variable analysis, each study contributed more than one effect size because they were treated separately.

## **Outliers**

A closer look at the effects sizes in Tables 3 and 4 revealed that the effect sizes from Yang (2001), Cirocki (2009), and Bell (2001) were larger than 3

Table 3. *Descriptive Statistics (Group Contrasts)*

	N (total)	n exp	n c	Pretest experimental		Pretest control		Posttest experimental		Posttest control		<i>d</i>
				<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Reading Speed												
Al- Homoud & Schmitt (2009)	65	43	22	60.08	19.41	61.62	21.61	93.57	21.18	87.75	28.75	0.23
Beglar et al. (2012)	52	35	17	103.09	25.14	87.54	12.42	119.93	25.16	90.51	14.10	1.44
Bell (2001)	26	14	12	68.10	14.37	78.45	15.39	127.53	29.31	92.54	19.72	1.40
Cha (2009)	20	10	10	97.80	33.04	77.00	32.36	178.40	69.91	99.60	33.84	1.44
Lao & Krashen (2000)	130	91	39	52.60	13.37	40.08	8.83	76.60	14.50	42.44	8.05	2.91
Mason & Krashen (1997) study 3	70	34	36	40.12	8.01	28.89	5.31	18.32	3.67	24.06	5.76	1.19
Matsui & Noro (2010)	122	60	62	88.57	29.71	89.29	24.54	108.32	33.89	94.79	26.90	0.44
Rob & Susser (1989)	125	63	62	—	—	—	—	—	—	—	—	0.96
Tanaka & Stapleton (2007)	190	96	94	28.84	5.81	29.96	8.09	36.00	8.28	32.41	8.17	0.44
Reading comprehension												
Al- Homoud & Schmitt (2009)	70	47	23	6.21	2.82	5.57	2.50	7.45	2.95	7.00	2.00	0.18
Beglar et al. (2012)	52	35	17	27.54	2.44	25.65	2.64	28.34	2.47	27.06	2.36	0.53
Bell (2001)	25	14	11	9.21	1.63	10.00	2.24	17.78	1.19	13.18	1.60	3.26
Cha (2009)	20	10	10	1.60	0.70	1.40	0.52	3.90	2.60	1.80	1.03	1.06
de Morgado (2009)	60	30	30	7.46	0.78	7.56	1.00	7.92	0.76	7.59	1.22	0.32
Lituanas et al. (2001)	60	30	30	7.61	5.19	7.89	4.81	32.57	14.80	12.28	5.77	1.81
Lituanas et al. (2001) <sup>a</sup>	60	30	30	2.36	0.83	2.55	0.73	5.25	1.16	3.96	0.88	1.25
Mason & Krashen (1997) study1	40	20	20	22.55	11.54	29.70	8.23	31.40	11.43	33.05	8.24	-0.17
Mason & Krashen (1997) study 2 <sup>b</sup>	79	40	39	29.55	8.87	31.30	11.04	48.08	8.86	41.88	11.50	0.60
Mason & Krashen (1997) study 2 <sup>c</sup>	49	31	18	16.74	8.00	17.56	7.42	33.71	9.02	25.69	10.15	0.84
Mason & Krashen (1997) study 3	78	40	38	29.45	8.31	30.13	8.07	45.52	8.28	44.29	9.16	0.14
Matsui & Noro (2010)	122	60	62	16.65	5.02	16.85	4.75	18.18	4.09	17.84	3.85	0.09
Nakanishi & Ueda (2011)	44	20	24	29.30	8.27	32.88	5.64	36.65	7.39	39.71	7.52	-0.41
Rezaee & Nourzadeh (2011)	51	26	25	9.00	2.83	8.59	3.13	10.80	2.95	9.08	2.36	0.64
Rob & Susser (1989)	125	63	62	—	—	—	—	—	—	—	—	0.28
Smith (2006)	102	51	51	22.30	8.00	23.90	8.50	36.70	7.80	34.80	7.90	0.24
Tanaka & Stapleton (2007)	190	96	94	22.01	5.50	21.23	6.22	24.35	6.77	21.83	7.17	0.36
Yang (2001) (Classes A & C)	61	31	30	62.30	1.68	61.90	1.43	75.30	1.35	66.70	0.99	7.26
Yang (2001) (Classes B & D)	59	29	30	62.10	1.94	61.30	1.65	73.70	1.15	67.00	1.35	5.34

Table 3 (continues).

Table 3. (continued).

	N (total)	N exp	N c	Pretest experimental		Pretest control		Posttest experimental		Posttest control		<i>d</i>
				<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Vocabulary												
Al-Homoud & Schmitt (2009)	65	45	20	19.11	6.10	14.45	5.03	24.96	9.56	21.35	7.01	0.43
Cha (2009)	20	10	10	36.00	30.24	28.20	14.05	42.20	32.19	30.90	14.52	0.45
Lao & Krashen (2000)	130	91	39	45.20	9.26	33.59	7.43	51.00	10.50	34.41	7.24	1.84
Lee (2007) study 1	103	65	38	37.90	12.00	44.00	13.00	44.00	11.50	45.00	12.60	-0.08
Lee (2007) study 1 <sup>d</sup>	103	65	38	37.90	12.00	46.30	11.40	44.00	11.50	51.30	10.40	-0.67
Lee (2007) study 2	206	67	139	47.10	10.00	46.90	10.10	52.10	8.20	51.80	9.80	0.03
Lee (2007) study 3	180	41	139	44.40	8.20	46.90	10.10	58.90	7.90	51.80	9.80	0.80
Rodrigo et al. (2004)	16	12	4	25.33	17.27	30.87	13.82	36.83	19.10	33.50	14.27	0.20
Tsang (1996)	96	48	48	61.77	8.42	61.89	9.48	70.56	4.63	67.40	5.83	-0.60
Yamamoto (2011)	67	33	34	64.64	8.98	60.15	13.14	65.42	7.88	61.97	9.29	0.40

Note. exp = experiment group; c = control group.

a The same participants took a different test.

b University student data are used.

c Junior college student data are used.

d The same participants took a different test.

Table 4. *Descriptive Statistics (Pre-Post Contrasts)*

	N exp	Pre exp		Post exp		<i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Reading Speed						
Cirocki (2009)	20	440.95	8.70	517.98	21.93	4.62
Fujita & Noro (2009)	51	75.05	24.22	85.05	25.15	0.41
Greenberg et al. (2006)	27	40.33	11.66	42.60	13.43	0.18
Iwahori (2008)	33	84.18	28.76	112.82	29.39	0.99
Lai (1993) S1	86	165.00	59.00	226.00	132.00	0.60
Lai (1993) S2	88	85.00	46.00	181.00	45.00	2.11
Lai (1993) S3	33	106.00	79.00	121.00	81.00	0.19
Taguchi et al. (2004)	10	80.88	19.14	108.24	32.88	1.02
Reading comprehension						
Fujimori (2006)	114	7.65	2.54	9.10	2.47	0.58
Fujita & Noro (2009)	68	9.15	3.43	10.01	3.89	0.23
Greenberg et al. (2006)	27	26.15	4.77	26.41	4.42	0.06
Hayashi (1999)	40	46.30	4.30	51.00	5.90	0.91
Iwahori (2008)	33	47.58	11.06	51.00	11.50	0.30
Lai (1993) S1	126	44.20	19.30	47.90	19.10	0.19
Lai (1993) S2	88	47.30	19.90	58.10	21.20	0.53
Lai (1993) S3	52	40.00	18.00	41.10	20.30	0.06
Lin (2010) male students	38	69.37	8.55	71.05	10.01	0.18
Lin (2010) female students	40	69.50	9.19	76.00	8.44	0.74
Mason (2003)	28	121.78	25.06	162.32	53.99	0.96
Mason (2007) English majors	16	22.40	8.80	29.40	11.70	0.68
Mason (2007) Non-English majors	24	14.90	5.30	19.30	7.00	0.72
Taguchi et al. (2004)	10	1.91	2.51	10.10	2.38	3.35
Takase (2007)	216	35.07	5.70	41.74	6.56	1.09
Takase (2009)	36	25.40	7.22	30.00	6.88	0.65
Yamashita (2008)	31	34.58	5.92	36.61	7.12	0.31
Yamashita (2008) <sup>e</sup>	31	37.73	10.25	43.05	10.24	0.52
Vocabulary						
Greenberg et al. (2006)	27	135.23	20.75	135.69	15.30	0.03
Hayashi (1999)	40	49.50	2.80	52.00	2.80	0.89
Horst (2005)	17	41.35	5.38	47.94	1.89	1.63
Horst (2005) <sup>f</sup>	17	33.80	8.18	43.59	4.30	1.50
Kweon & Kim (2008)	12	128.75	40.23	214.91	33.17	2.34
Zimmerman (1997)	17	138.82	19.08	153.71	13.89	0.90

*Note.* exp = experiment group.

<sup>e</sup> The same participants took a different test.

<sup>f</sup> The same participants took a different test.

( $d = 7.26, 5.34, 4.62,$  and  $3.26,$  respectively), which are statistical outliers (Plonsky & Gass, 2011); therefore, the four effect sizes were excluded from the analysis. Cooper (2010) also argued that “meta-analysts should always be cautious in interpreting any effect size based on a small number of data points. When samples are small, a single extreme value can create an exceptionally large effect size estimate” (p. 173). One of the effect sizes by Taguchi et al. (2004) was larger than 3.0, but was retained because when averaged it became less than 3.0.

Tables 5 and 6 show the remaining effect sizes included in the analysis. The 34 studies provided 43 unique effect sizes (22 effect sizes for group contrasts and 21 effect sizes for pre-post contrasts, respectively) and a total sample size of 3,942 participants. The sample sizes of the group contrasts ranged from 10 to 96 ( $M = 42.32$ ) for the experimental groups and 10 to 139 ( $M = 38.95$ ) for the control groups. This finding indicates that the researchers attempted to have similar numbers of students in each group. On the other hand, the sample sizes of the pre-post contrasts ranged from 10 to 216 ( $M = 51.29$ ); this amount of variance suggests that it is relatively easy to conduct research with one group of participants. The  $d$  values were calculated manually using

Excel. Once all the effect sizes for the two types of contrasts were calculated, they were entered into the Comprehensive Meta-Analysis software (version 2.2, Biostat, 2007) and the analysis in which all the effect sizes were weighted according to sample sizes was conducted.

Table 5. *Effect Sizes Aggregated for Calculating the Overall Effect of Extensive Reading (Group Contrasts)*

	N (total)	N exp	N c	<i>d</i>
Al- Homoud & Schmitt (2009)	70	47	23	0.18
Beglar et al. (2012) <sup>a</sup>	52	35	17	0.99*
Bell (2001)	26	14	12	1.40
Cha (2009)	20	10	10	1.06
de Morgado (2009)	60	30	30	0.32
Lao & Krashen (2000)	130	91	39	2.38*
Lee (2007) study 1	103	65	38	-0.08
Lee (2007) study 2	206	67	139	0.03
Lituanas et al. (2001)	60	30	30	1.53*
Mason & Krashen (1997) 1	40	20	20	-0.17
Mason & Krashen (1997) 2 <sup>b</sup>	79	40	39	0.60
Mason & Krashen (1997) 2 <sup>c</sup>	49	31	18	0.84
Mason & Krashen (1997) 3	78	40	38	0.14
Matsui & Noro (2010)	122	60	62	0.09
Nakanishi & Ueda (2011) <sup>d</sup>	44	22	24	-0.41
Rezaee & Nourzadeh (2011)	51	26	25	0.64
Rob & Susser (1989) <sup>e</sup>	125	63	62	0.96
Rodrigo et al. (2004)	16	12	4	0.20
Smith (2006)	102	51	51	0.24
Tanaka & Stapleton (2007)	190	96	94	0.36
Tsang (1996)	96	48	48	-0.60
Yamamoto (2011) <sup>f</sup>	67	33	34	0.40

*Note.* Key: exp = experiment group; c = control group. An asterisk indicates an averaged effect size within the study.

a Out of the three experimental groups, group 3 was chosen because it has the largest sample size.

b Data of university students are used.

c Data of junior college students are used.

d Control 2 was chosen because it has the largest sample size.

e The effect size was calculated with the sample size and  $F = 28.921$ .

f Total scores of the Vocabulary Levels Tests were used.

Table 6. *Effect Sizes Aggregated for Calculating the Overall Effect of Extensive Reading (Pre-Post Contrasts)*

	N exp	<i>d</i>
Fujimori (2006)	114	0.58
Fujita & Noro (2009)	68	0.23
Greenberg et al. (2006)	27	0.03
Hayashi (1999) Intermediate level score	40	0.91
Hayashi (1999) Beginning level score	36	1.68
Horst (2005)	17	1.57*
Iwahori (2008)	33	0.30
Kweon & Kim (2008) <sup>g</sup>	12	2.34
Lai (1993) S1 <sup>h</sup>	126	0.19
Lai (1993) S2 <sup>i</sup>	88	0.53
Lai (1993) S3 <sup>j</sup>	52	0.06
Lin (2010) Male students	38	0.18
Lin (2010) Female students	40	0.74
Mason (2003) <sup>k</sup>	36	1.92
Mason (2007) English majors	16	0.68
Mason (2007) Non-English majors	24	0.72
Taguchi et al. (2004)	10	2.19*
Takase (2007)	216	1.09
Takase (2009) <sup>l</sup>	36	0.65
Yamashita (2008)	31	0.41*
Zimmerman (1997)	17	0.90

*Note.* Key: exp = experiment group. An asterisk indicates an averaged effect size within the study.

g Mean scores for noun were used.

h Grade 7 students. Data of reading comprehension tests are used.

i Grade 8 students. Data of reading comprehension tests are used.

j Grade 9 students. Data of reading comprehension tests are used.

k A group that wrote summaries in English and received feedback was chosen because of the largest sample size.

l Group 3 is chosen because it has the largest sample and students read more books than the rest of the groups.

## Coding the Primary Studies

All 34 studies were coded as specified by the coding table displayed in

Table 7. The coding table was constructed by reviewing previously published

meta-analyses. The table includes nine variables: sample size, participant,

length of instruction, control group, areas of interest, test format, test use, test reliability, environment, number of books read, and number of words read. The results of the coded variables across studies are shown in Table 8. Length of instruction was a continuous variable, but following Plonsky (2011), it was divided into five dichotomous categories. This approach was selected because most of the studies were conducted over one or two semesters.

Coding is subject to error; thus, it is crucial to assess the degree of agreement among coders (Cooper & Hedges, 1994). To establish the reliability of the coding procedures used in this meta-analysis, I consulted with a researcher who is familiar with meta-analysis and who has published related studies (In'nami & Koizumi, 2009, 2010). He and I coded ten randomly selected studies. Intercoder reliability was calculated using Cohen's kappa ( $\kappa$ ) coefficient, which reflects both agreement and error. The intercoder reliability for coding the ten primary studies was acceptable at  $\kappa = .969$ . Disagreements were discussed, and the remaining studies were coded by the author.

## **Rationale for Each Moderator**

In this section, the rationale for including each moderator is described.

This aspect of the study has often been neglected in previous studies (See Ahn et al. [2012] for a discussion of this issue).

***Participant.*** Extensive reading was used with multiple age groups. Investigating the effect of extensive reading on learners' of different ages will provide information concerning who benefits the most from extensive reading instruction. In addition, this analysis will also show how many studies have been conducted with each age group.

***Length of instruction.*** Length of instruction is a crucial aspect of extensive reading. Information about the effects of differing lengths of instruction is valuable for instructors when creating a reading syllabus. The weakness of using time as an independent variable, however, is that the amount of reading, not time, is the key to gains in extensive reading studies. For instance, some students can read more in one semester than others read in one year. In other words, "seat time" is not a reliable measure or guarantee of actual effort or

learning. This is why extensive reading researchers should indicate the number of running words or standard words students read.

***Control group.*** The inclusion of a control group is of the utmost importance when investigating the effect of any treatment, as the control group can show the degree to which the outcome variable(e) are influenced by aspects of the curriculum other than the treatment.

***Areas of interest.*** Researchers focus on specific skills when conducting extensive reading research. In this study, four main areas emerged: reading speed, reading comprehension, vocabulary, and grammar.

***Test used.*** The tests used are closely related to the areas of interest. Three types of tests were used in the studies included in this meta-analysis: cloze tests, reading comprehension tests, and reading speed assessments.

***Test use and reliability.*** It is important to note the reliability estimate for each quantitative instrument. Highly reliable tests provide more reliable insights than those with low reliability.

***Environment, number of books read, and number of words read.*** This information describes where the study was conducted, and how many books and running words the participants read during the treatment.

### **Publication Bias (Sensitivity Analysis)**

Assessing publication bias is a crucial element of a meta-analysis, yet few meta-analysis studies in the field of second language acquisition have reported any information concerning this issue (e.g., Li, 2010; Norris & Ortega, 2000).

Checking for publication bias is important because the results obtained in studies in which publication bias was not assessed can be overturned later by new meta-analyses in which it was assessed (Ellis, 2010). There are two major ways to assess publication bias, the fail-safe  $N$  and the funnel plot. The fail-safe  $N$  involves computing how many nonsignificant studies are needed to create a

Table 7. *Coding Table*

Variables	Codes
A. Sample size	
B. Participant	1 Junior high school students 2 high school students 3 university students (Including junior college) 4 adults 5 children (elementary school or below)
C. Length of instruction	1 one semester (less than 3 months) 2 two semesters (from 3 months to 6 months) 3 one year (from more than 6 months to one year) 4 one year to two years 5 two years to three years
D. Control group	1 none 2 one group 3 two groups or more
E. Areas of interest	1 reading speed 2 reading comprehension 3 vocabulary 4 grammar 5 other
F. Test used	1 cloze test 2 reading comprehension 3 reading speed 4 other
G. Test use	1 Same test version used in pre and post 2 Parallel version
H. Test reliability	1 Not reported 2 Reported (data based on the current study) 3 Reported (data cited from the test manual)
I. Environment	1 ESL 2 EFL 3 Native language
J. Number of books read	1 Not specified Other (Actual number if specified)
K. Number of words read	1 Not specified Other (Actual number if specified)

Table 8. *Coded Variables Across Studies*

	A	B	C	D	E	F	G	H	I	J	K
Al- Homoud & Schmitt (2009)	70	3	1	2	1 & 2 & 3	2 & 3	1	1	2	1	1
Beglar et al. (2012)	97	3	3	2	1 & 2	2 & 3	1	2	2	24.36	20,0170
Bell (2001)	26	4	2	2	1 & 2	2 & 3	1	1	2	1	1
Cha (2009)	20	2	1	2	1 & 2 & 3	2 & 3	1	1	2	1	1
de Morgado (2009)	60	3	1	2	2	2	2	1	2	1	1
Fujimori (2006)	114	2	3	2	2 & 5	2 & 4	2	1	2	1	1
Fujita & Noro (2009)	68	2	2	1	1 & 2	2 & 3	2	2	2	1	2,517
Greenberg et al. (2006)	27	4	2	1	1 & 2 & 3	1 & 2 & 3	1	3	3	1	1
Hayashi (1999)	40	3	3	1	2 & 3	2	1	1	2	5.30	1
Horst (2005)	17	4	1	1	3	4	*	1	1	10.52	1
Iwahori (2008)	33	2	1	1	1 & 2	1 & 3	1	2	2	4	1
Kweon & Kim (2008)	12	3	1	1	3	4	1	1	2	3	13,4013
Lai (1993) S1	126	1	1	1	1 & 2	2 & 3	2	1	2	16	1
Lai (1993) S2	88	1	1	1	1 & 2	2 & 3	2	1	2	18.50	1
Lai (1993) S3	52	1	1	1	1 & 2	2 & 3	2	1	2	14.20	1
Lao & Krashen (2000)	130	3	1	2	1 & 3	3 & 4	1 & 2	3	2	6	388,000
Lee (2007) 1a	103	3	1	3	3	1 & 4	1	1	2	1	1
Lee (2007) 1b	103	3	1	3	3	1 & 4	1	1	2	1	1
Lee (2007) 2	206	3	3	3	3	1 & 4	1	1	2	1	1
Lee (2007) 3	180	3	3	3	3	1 & 4	1	1	2	1	1
Lin (2010)	78	2	1	1	2	2	1	1	2	3	1
Lituanas et al. (2001)	60	1 & 2	2	3	2	2	1	1	2	1	1
Mason (2003)	104	3	4	1	2 & 5	1, 2, & 4	1	2	2	1	500,000
Mason (2007)	40	3	1	1	2 & 5	1	1	2	2	1	1
Mason & Krashen (1997) 1	40	3	1	2	2	1	1	2	2	30	1
Mason & Krashen (1997) 2a	79	3	3	1	2	1	1	2	2	1	1
Mason & Krashen (1997) 2b	49	3	3	1	2	1	1	2	2	1	1
Mason & Krashen (1997) 3a	78	3	3	2	1 & 2	1, 2, & 3	1	2	2	1	1
Mason & Krashen (1997) 3b	74	3	3	2	1 & 2	1, 2, & 3	1	2	2	1	1
Matsui & Noro (2010)	122	1	3	2	1 & 2	2 & 3	2	2	2	1	18,907

Table 8 (continues).

Table 8. (continued)

	A	B	C	D	E	F	G	H	I	J	K
Nakanishi & Ueda (2011)	89	3	3	3	2	2	2	3	2	147	73,646
Rezaee & Nourzadeh (2011)	51	4	1	2	2	2	1	2	2	4	1
Rob & Susser (1989)	125	3	3	3	1 & 2	2 & 3	2	1	2	1	1
Rodrigo et al. (2004)	27	3	1	3	3, 4, & 5	1 & 4	1	1	2	1 ~ 5	1
Smith (2006)	102	2	3	2	2	1 & 2	2	3	2	1	1
Taguchi et al. (2004)	20	3	2	2	1	3	2	1	2	3 ~ 6	1
Takase (2007)	216	2	3	1	2	2	1	1	2	1	71,653
Takase (2009)	138	3	3	1	2	2	1	1	2	158.6	188,177
Tanaka & Stapleton (2007)	190	2	2	3	1 & 2	2 & 3	2	1	2	1	1
Tsang (1996)	96	1 & 2	2	2	3 & 5	4	1	1	2	8	1
Yamamoto (2011)	67	3	1	2	3	4	1	2	2	1	1
Yamashita (2008)	31	3	1	1	2 & 5	1 & 2	2	2	2	11	1
Zimmerman (1997)	35	3	1	2	3	4	1	2	1	1	1

Note. Only studies with different participants are coded.

\*The participants took the same pretest, but post-tests were created individually based on the books a participant had chosen to read.

nonsignificant effect; however, there are criticisms over the measure.

Borenstein, Hedges, Higgins, and Rothstein (2009) claimed that “this approach is not generally appropriate for analyses that focus on effect sizes” (p. 285).

The alternative approach is the funnel plot, which displays the relationship between sample size and effect size. It is primarily used as a visual aid to assess publication bias. A well-balanced data set forms a symmetric funnel shape, in which publication bias is unlikely; an unbalanced data set forms an asymmetric funnel that indicates the possibility of publication bias. To get better sense of publication bias, it is used to assess publication bias for each contrast. In the funnel plots (see Figures 1 and 2), sample sizes are plotted on the  $y$ -axis and effect sizes for each study on the  $x$ -axis. Figure 1 shows the funnel plot for group contrasts and Figure 2 displays the funnel plot for pre-post contrasts. Large studies tend to appear toward the top of the graph and cluster around the mean effect size whereas smaller studies are located around the bottom of the graph. The smaller studies create variance; thus, they are scattered across a range of values. In the presence of publication bias, plots become asymmetric on either side of the mean.

The funnel plots for the two contrasts indicate the following patterns.

First, there are generally only a few studies with large samples at the top of the graph. Second, many studies with smaller sample sizes are dispersed symmetrically around the mean at the bottom of the graph. For the pre-post contrasts, however, there should be more positive small *N*-size studies. In a sense, this is understandable, given the fact that the pre-post contrasts did not include control groups.

### **Data Summary**

The 34 studies provided 43 unique effect sizes (22 effect sizes for group contrasts and 21 effect sizes for pre-post contrasts, respectively). These effect sizes were used for calculating the overall effectiveness of extensive reading. Initially, 66 effect sizes, excluding four effect sizes from Yang (2001), Cirocki (2009), and Bell (2001), were employed for the moderator variable analyses. There used to be few studies of the effects of extensive reading on reading speed; however, as noted in the Methods chapter, there are now a similar number of studies of reading speed and vocabulary acquisition; 17 studies were focused on reading speed and 15 studies on vocabulary acquisition. Reading

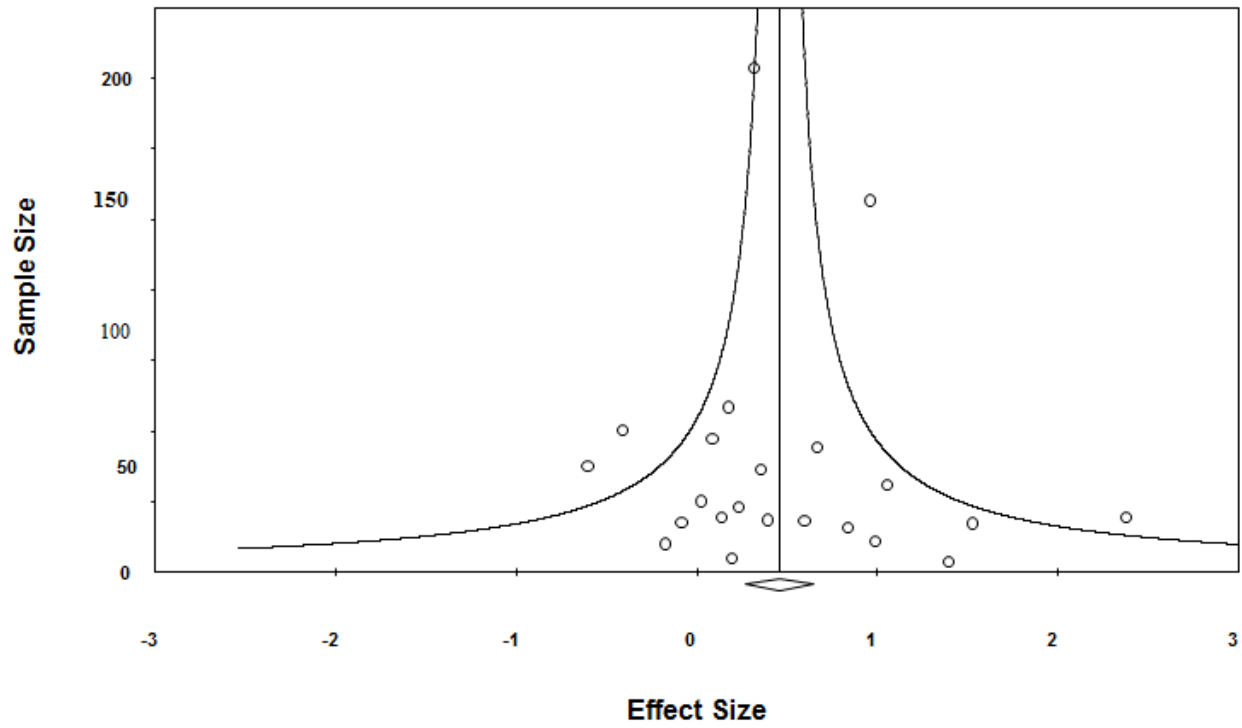


Figure 1. Funnel plot of group contrasts.

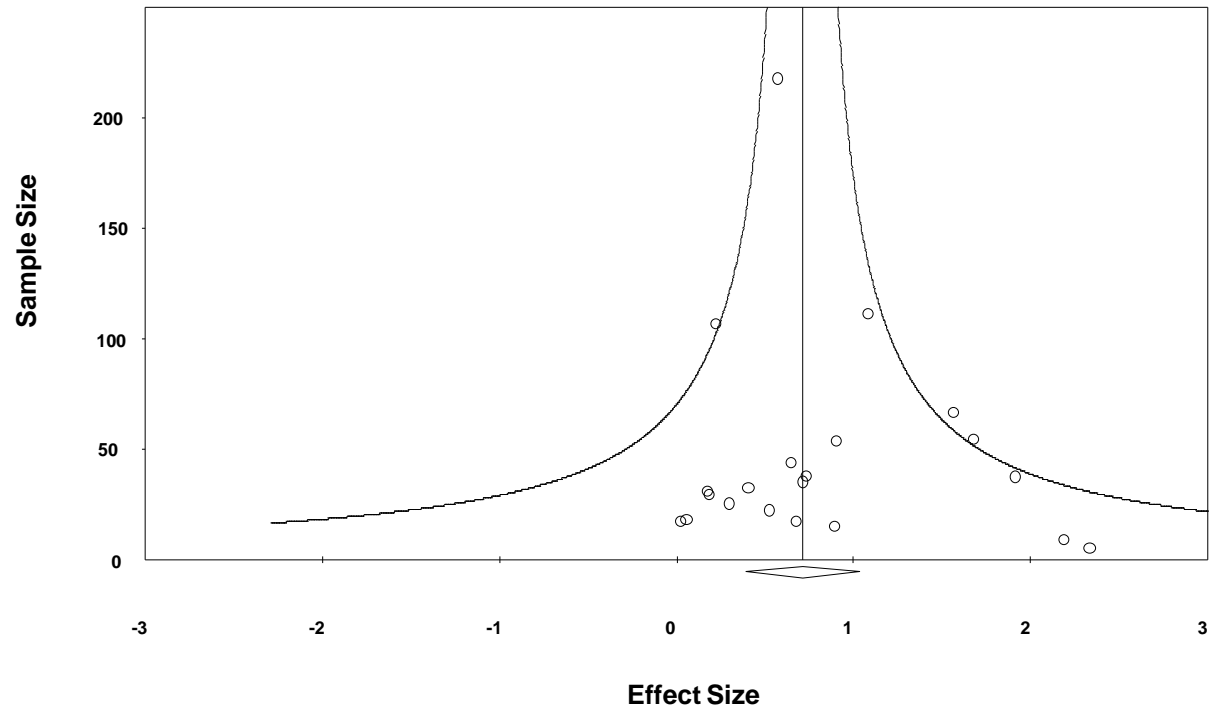


Figure 2. Funnel plot of pre-post contrasts.

Comprehension was assessed in 31 studies. That the largest number of studies concerned reading comprehension is understandable because comprehension is the purpose of reading.

## **CHAPTER 5**

### **RESULTS**

In this chapter, I describe the results of the analyses and compare several research features based on the coded variables across studies to provide a fuller picture of extensive reading research. First, the year of the publication of extensive reading studies and research features of those studies are described. Next, the overall effects of extensive reading for the group contrasts and pre-post contrasts are reported. Finally, moderator variables for each subgroup are examined.

#### **The Research Synthesis**

##### **Publication Dates**

In this section, first, the year of publication for extensive reading research is discussed, as it is commonly used in interpreting results of meta-analysis (Greenhouse & Iyengar, 1994; Lipsey & Wilson, 2001). Table 9 summarizes the publication dates of the studies included in Appendices 1, 2, 3, which are summaries of extensive reading research by Day and Bamford (1998),

Horst (2005), and Iwahori (2008), and the studies included in this meta-analysis. These studies are compared to the studies in this meta-analysis to describe the distribution of the studies. Table 9 shows empirical studies of extensive reading that have appeared since 1981. Approximately 50% of the studies were published between 1981 and 1995; however, many of the researchers did not report enough statistical information for their study to be included in this meta-analysis. For this reason, only two studies published between 1981 and 1995 were included in the meta-analysis, while 32 studies published since 1996 were included. This finding indicates that recent researchers have provided the descriptive statistics needed to calculate effect sizes. This is a positive tendency that journal reviewers, journal editors, and researchers need to be aware of and continue. Considering the information in Appendices A, B, C, and the number of extensive reading studies that have been published since the 1980s, a more extensive meta-analysis on extensive reading could be conducted in the near future to see how the field has developed. This could be done in a manner similar to Doughty's (2003) reanalysis of the data in Norris and Ortega's (2000) meta-analysis.

Table 9. *Publication Dates of Studies Included in Appendices A, B, C, and in this Meta-Analysis*

Year of publication	Summary of Appendices A, B, C	Studies in this meta-analysis
2001-2012	4	27
1996-2000	6	5
1991-1995	4	1
1981-1990	7	1

*Note.* The number in the summary of Appendices 1, 2, 3 also includes Horst (2005) and Iwahori (2008).

### **Research Features**

Table 10 illustrates the research features of studies included in this meta-analysis and the percentages of each feature. University students (56%) were the majority of the participants in ER studies, followed by high school students (22%), and junior high school students (13%).

In terms of length of instruction, 47% of the studies were short-term, which is defined as less than three months in this study. This percentage increases to 63% if the length of extensive reading instruction is extended to six months. Only 15 studies were conducted for between six months and one year. One study, Mason's (2003) dissertation, was conducted for more than one year. These findings clearly indicate that longitudinal studies of extensive reading conducted for more than one year are still rare, and greatly needed. It should be

noted, however, that the primary studies differed substantially in terms of the timing of the posttests.

More than half of the studies (61%) included one or more control groups, but this number is low, considering the importance of control groups in experimental and quasi-experimental research. Without a control group, there is no way to determine the degree to which other aspects of the curriculum influenced the results of the treatment.

Three main areas of research interests for extensive reading researchers were gains in reading comprehension (74%), increases in reading speed (40%), and vocabulary acquisition (35%). The studies reviewed here indicated that extensive reading resulted in improvements in all three areas. To investigate these areas, the researchers administered proficiency tests with reading comprehension subsections, such as TOEFL and STEP EIKEN, and the Edinburgh project of extensive reading (EPER) placement/progress tests (60%). These tests were followed by reading rate tests (40%) and cloze tests (37%).

Most researchers choose to administer the same test two or more times (67%) rather than a parallel version of the test (33%). This practice opens the door to test-retest effects. Given the fact that 63% of studies had a treatment

period of less than six months, the test-retest effects might have influenced the outcomes. An additional concern is that more than half of the researchers (56%) did not report the reliability of the tests they used, and 9% reported reliability coefficients listed in the test manual. This is unacceptable because the reliability of a test can differ significantly from one sample to another because reliability results from an interaction of a particular sample of persons with the test items. The issue of reliability is important because the results of these studies are questionable to the degree that the tests were unreliable. Test reliability based on data gathered in the study was reported by only 35% of the researchers.

Finally, 93% of the studies were conducted in EFL contexts and 5% (two studies) in an ESL context. Information concerning the participants and what tests were utilized in each study is summarized in Table 13.

### **Meta-Analysis**

Prior to moving into the results of the analysis, the interpretation of effect sizes should be shown here again to provide an overview of discussion of effect sizes. Table 11 shows the criteria presented by Cohen (1988) and Oswald

Table 10. *Research Features*

Features		N of studies	%
Participant	1 Junior high school students	6	13
	2 high school students	10	22
	3 university students (Including junior college)	25	56
	4 adults	4	9
	5 children (elementary school or below)	0	0
Length of instruction	1 one semester (less than 3 months)	20	47
	2 two semesters (3 to 6 months)	7	16
	3 one year (> 6 months to one year)	15	35
	4 one year to two years	1	2
	5 two years to three years	0	0
Control group	1 none (i.e., pre-post design)	17	40
	2 one group	17	40
	3 two groups or more	9	21
Areas of interest	1 reading speed	17	40
	2 reading comprehension	32	74
	3 vocabulary	15	35
	4 grammar	1	2
	5 other	5	12
Test used	1 cloze test	16	37
	2 reading comprehension	26	60
	3 reading speed	17	40
	4 other	13	30
Test used	1 Same test used for pre- and posttest	29	67
	2 Parallel version	14	33
Test reliability	1 Not reported	24	56
	2 Reported (data based on the current study)	15	35
	3 Reported (data cited from the test manual)	4	9
Environment	1 ESL	2	5
	2 EFL	40	93
	3 Native language	1	2

*Note.*  $k = 43$ . (See Table 8 for details). All the percentages are rounded, thus some of them do not add up to 100. The total number of participants was 45 because two studies included two types of participants. The percentages of areas of interest and test used do not add up to 100 because some studies utilized multiple tests.

and Plonsky (2010) concerning the interpretation of effect sizes. Oswald and

Plonsky's more conservative criteria were used in this study.

Table 11. *Summary of the Interpretation of Effect Sizes*

	Cohen (1988)	Oswald and Plonsky (2010)
Small effect	$d = .20$	$d = .40$
Medium effect	$d = .50$	$d = .70$
Large effect	$d = .80$	$d = 1.00$

As stated above, this meta-analysis included 34 studies that provided 43 unique effect sizes (22 effect sizes for group contrasts and 21 effect sizes for pre-post contrasts, respectively) and a total sample size of 3,942 participants. In order to address the overall effectiveness of extensive reading, the overall random-effects effect size, Cohen's  $d$ , of the effects of extensive reading on L2 reading proficiency was examined. The results of the group contrasts, including effect sizes, number of studies, sample size, and confidence intervals, are shown in Table 12. The results indicated that there is a small effect of extensive reading on L2 reading proficiency ( $d = .46$ ), that is, experimental groups that received an extensive reading treatment performed .46 standard deviation units better than the control groups. This effect size was significantly greater than no effect because the confidence interval did not include 0. This differed from the confidence interval for the random-effects effect size in Yang's study (2001), which included 0 (i.e., not significantly greater than no effect). The homogeneity test was not statistically significant ( $Q = 18.37, df = 21, p = .63$ ),

but when including Yang's studies (2001), it was statistically significant ( $Q = 1003.49$ ,  $df = 23$ ,  $p = .00$ ), which means the two studies created a degree of heterogeneity.

The effectiveness of extensive reading for pre-post contrasts was also investigated. The results showed a medium effect ( $d = .71$ ) for extensive reading (See Table 13). The homogeneity test was not statistically significant ( $Q = 5.86$ ,  $df = 20$ ,  $p = .99$ ). Although the studies included here did not include a control group, the overall effect size was large enough to conclude that extensive reading exerted a positive instructional effect.

### **Moderator Variable Analysis**

In this section, moderator variables for each contrast were examined. This study was primarily focused on revealing systematic disparity in the effectiveness of extensive reading across age groups and length of instruction. To address these issues, effect sizes were calculated for coded-subgroups (see also Tables 12 and 13). To provide comprehensive answers to the research

Table 12. Overall Meta-Analysis Results for Group Contrasts

Group	Subgroup	<i>d</i>	<i>k</i>	N	CI (Low, High)
Overall		0.46	22	1,788	[0.27, 0.65]
Participant	1 Junior high school students	-0.05	3	278	[-0.83, 0.74]
	2 high school students	0.57	3	312	[-0.36, 1.50]
	3 university students (Including junior college)	0.48	14	1,121	[0.22, 0.74]
	4 adults	0.67	2	77	[-0.44, 1.79]
	5 children (elementary school or below)	—	—	—	—
Length of instruction	1 one semester (less than 3 months)	0.36	9	557	[0.11, 0.61]
	2 two semesters (from 3 months to 6 months)	0.03	4	372	[-0.86, 0.93]
	3 one year (from more than 6 months to one year)	0.52	9	859	[0.09, 0.95]
	4 one year to two years	—	—	—	—
	5 two years to three years	—	—	—	—
Areas of interest	1 reading speed	0.98	9	800	[0.64, 1.33]
	2 reading comprehension	0.63	16	1,367	[0.06, 1.21]
	3 vocabulary	0.18	9	883	[-0.60, 0.96]
	4 grammar	—	—	—	—
	5 other	—	—	—	—
Test use	1 Same test version used in pre and post	0.31	15	1,013	[-0.17, 0.79]
	2 Parallel version	0.44	7	775	[-0.01, 0.88]
Test reliability	1 Not reported	0.49	11	972	[0.19, 0.79]
	2 Reported (data based on the current study)	0.39	8	538	[-0.27, 1.04]
	3 Reported (data cited from the test manual)	0.38	3	278	[-1.10, 1.87]

Note. The overall *d* indicates both fixed effect and random effect. When two of Yang (2001) studies were retained, *d* = 2.02 (FE) and 1.06 (RE), *df* = 23.

Table 13. Overall Meta-Analysis Results for Pre-Post Contrasts

Group	Subgroup	<i>d</i>	k	N	CI (Low, High)
Overall		0.71	21	2,154	[0.39, 1.04]
Participant	1 Junior high school students	0.27	3	532	[-2.13, 2.66]
	2 High school students	0.61	6	1,018	[0.25, 0.97]
	3 University students (Including junior college)	1.12	10	516	[0.23, 2.01]
	4 adults	1.48	2	88	[0.05, 2.90]
	5 children (elementary school or below)	—	—	—	—
Length of instruction	1 one semester (less than 3 months)	0.89	12	988	[-0.01, 1.79]
	2 two semesters (from 3 months to 6 months)	0.24	3	210	[-0.67, 1.15]
	3 one year (from more than 6 months to one year)	0.74	5	884	[0.36, 1.12]
	4 one year to two years	1.92	1	72	[-0.71, 4.55]
	5 two years to three years	—	—	—	—
Areas of interest	1 reading speed	0.61	7	328	[-2.79, 4.02]
	2 reading comprehension	0.72	17	987	[0.40, 1.05]
	3 vocabulary	1.25	5	113	[0.14, 2.35]
	4 grammar	—	—	—	—
	5 other	—	—	—	—
Test use	1 Same test version used in pre and post	1.05	13	571	[0.44, 1.65]
	2 Parallel version	0.58	8	506	[0.20, 0.96]
Test reliability	1 Not reported	0.77	13	825	[0.42, 1.13]
	2 Reported (data based on the current study)	0.45	7	225	[-0.33, 1.22]
	3 Reported (data cited from the test manual)	0.03	1	27	[-5.73, 5.79]

Note. The overall *d* indicates both fixed effect and random effect. When Cirocki (2009) was retained,  $d = .72$  (FE and RE),  $df = 21$ .

questions, confidence intervals were examined to identify any significant differences between groups. It should be noted, however, that all groups substantially overlapped with each other in terms of their 95% confidence intervals, indicating no statistically significant difference, for example, between the three categories of instructional length (one semester vs. two semesters vs. one year) (see Tables 12 and 13). Finally, the test usage for extensive reading research was investigated.

First, five age groups were investigated. As for group contrasts, due to the small number of studies, the confidence intervals for the effect sizes for junior high school students, high school students, and adults all included zero; however, the effect size for university students indicated a small effect ( $d = .48$ ) in which the confidence interval did not include zero, which indicates a statistically significant difference between the observed effect and the null hypothesis of no effect. Out of the 22 studies included in the group contrasts analysis, only three studies involved junior high school students, three involved high school students, 14 involved university students, and two involved adults. On the other hand, the effect size for junior high school students included zero for pre-post contrasts. Out of 21 studies included in the pre-post contrasts

analysis, only three studies involved junior high school students, six involved high school students, ten involved university students, and two involved adults. A medium effect for high school students ( $d = .61$ ) and large effects for university students ( $d = 1.12$ ) and adults ( $d = 1.48$ ) were obtained.

The other major moderator variable is the length of instruction. In an empirical framework, one can assume that longer treatments should result in greater gains due to longer time-on-task, but due to the constraints discussed above, 63% of the studies in this meta-analysis were administered for less than six months and only 16 studies were conducted for one year or longer. Out of 22 studies included for in the group contrasts analysis, nine studies were conducted over one year, four over two semesters, and nine over one semester. One semester of instruction produced a small effect of  $d = .36$ , the confidence interval for two semesters included zero, and the effect size for one year was a small to medium effect ( $d = .52$ ).

Three of the confidence intervals for the effect sizes for pre-post contrasts included zero, and the effect size for one year of reading produced a medium effect ( $d = .74$ ). The results partly supported the notion that longer treatment periods produced larger effect sizes.

With respect to group contrasts, a large effect for reading speed ( $d = .98$ ) and a medium effect for reading comprehension ( $d = .63$ ) were obtained. The two confidence intervals for test use included zero and studies in which test reliability was not reported produced a small effect ( $d = .49$ ).

Turning to pre-post contrasts, a medium effect for reading comprehension ( $d = .72$ ) and a large effect for vocabulary ( $d = 1.25$ ) were obtained. As opposed to the group contrasts, a large effect ( $d = 1.05$ ) was found in studies in which the same test was used two or more times and a small to medium effect ( $d = .58$ ) was found for studies in which a parallel version of the test was attained. Studies in which test reliability was not reported produced a medium effect ( $d = .77$ ).

Table 14 summarizes the information concerning the participants' L1 and tests used to assess improvement for each study. Out of 34 studies, 20 different tests were administered. In addition to these 20 tests, researchers utilized self-made reading speed tests or self-made cloze tests. These tests were made based on textbooks participants used in class.

Table 14. *Participants' L1 and Tests Used in Each Study*

L1	<i>n</i>	Tests used	<i>n</i>
		Edinburgh Project on Extensive Reading (EPER)	
Japanese	17	Placemen/Progress Tests	5
Chinese	7	Vocabulary Levels Test (VLT)	4
Korean	2	TOEFL (Pre-TOEFL included)	3
Arabic	2	STEP EIKEN	2
Native speaker	2	The Secondary Language English Proficiency Test (SLEP)	2
Mixed (ESL class)	2	The Cambridge Preliminary English Test	1
		The Nation Vocabulary Test	1
Spanish	1	Woodcock Johnson Word Letter-Identification	1
Tagalog	1	Peabody Picture Vocabulary Test	1
		Boston Naming Test	1
		TOEIC	1
		Productive vocabulary levels test	1
		English Web VocabProfile	1
		The Informal Reading Inventory	1
		The Gray Standardized Oral Reading Test	1
		Burns/Roe Informal Reading Inventory	1
		Accuracy Level Test & Rate Level Test	1
		Multiple Skills Series Midway Placement Test	1
		The Standardized Reading Test Junior Version (Used three times by Lai [1993])	1
		College Students English Proficiency Test (CSEPT)	1

*Note.* If the same researcher conducted multiple studies, it is counted as a single study.  $N = 34$ .

The nationalities of the research participants were also investigated.

Japanese ( $n = 17$ ) and Chinese ( $n = 7$ ) made up the majority of the participants followed by Korean ( $n = 2$ ), Saudi Arabian ( $n = 2$ ), native speakers of English ( $n = 2$ ). One study was conducted in Venezuela and two were conducted in the Canadian ESL context. One study was conducted in Europe by Cirocki (2009), but it was not included in this meta-analysis because its effect size was an outlier.

A variety of assessment tools have been used in extensive reading research. The EPER placement/progress tests were used most frequently ( $n = 5$ ), and the Vocabulary Levels Test (Schmitt et al., 2001) was used most frequently by researchers who measured students' lexical knowledge ( $n = 4$ ); other vocabulary tests included the Peabody Picture Vocabulary Test, the Productive Vocabulary Levels Test, and English Web VocabProfile. The TOEFL reading section was used in three studies and the STEP EIKEN test was used in two studies. Some researchers referred to previous literature in support of their choice of instruments, but as indicated in Table 14, many disparate assessments were used, a finding that makes it difficult to contrast the results of different studies.

## CHAPTER 6

### DISCUSSION

#### Overall Effectiveness of Extensive Reading

The primary goal of this study was to address the magnitude of the overall effect of extensive reading on reading comprehension. The first research question asked about the effectiveness of extensive reading instruction. It was found that students who received extensive reading instruction outperformed students who did not. The effect size for group contrasts ( $d = .46$ ) and for pre-post contrasts ( $d = .71$ ) was small, and considerably smaller than Norris and Ortega's (2000) findings for the overall effects of L2 instruction ( $d = .96$ ) as well as the effects of implicit focus-on-form treatments ( $d = .69$ ). The effect size for the pre-post contrasts, however, needs to be interpreted cautiously because (a) no control groups were included in these studies and (b) there might have been a test-retest effect given that the same tests were administered twice. An additional concern is the possibility of regression to the mean. Researchers need to conduct tests at least three times to avoid this possibility (e.g., see Nakanishi and Ueda (2011) for an example). Furthermore, the effect sizes ( $d$ ) of studies in

the pre-post contrasts analysis ranged from .03 to 2.34 (see Table 6). The lack of a negative effect size might arise because of the lack of control groups. This finding is in contrast to the group contrasts, in which four studies produced negative effect sizes ranging from -0.60 to -0.08 (see Table 5).

### **Learner's Age**

The second research question asked whether the learners' ages impact learning differently. The results showed that three out of the four categories in the group contrasts included zero in the confidence interval; the one significant finding was that the effect size for university students was small ( $d = .48$ ).

These results are, for the most part, due to the small number of studies included in the analysis; junior high school and high school students participated in three studies each and adults participated in two studies.

With respect to pre-post contrasts, the confidence interval for the youngest group (i.e., junior high school students) included zero. A medium effect ( $d = .61$ ) was found for high school students, a large effect ( $d = 1.12$ ) was found for university students, and a large effect ( $d = 1.48$ ) was found for adults. Although all groups substantially overlapped with each other in terms of their

confidence intervals, which indicates that there were no statistically significant differences, the results suggest that the effect of extensive reading might increase with older participants. Extensive reading might be more beneficial for older learners who have learned the foreign language explicitly, as it might lead them to draw on and proceduralize their explicit knowledge. This possibility stands in contrast to younger learners, who are still acquiring knowledge about the foreign language. Another factor concerns the maturity of the participants in terms of their cognitive processing. As individuals age, they are able to understand and process more complex information, a development that could lead them to read more. These results are similar to the findings of the National Reading Panel (NRP) report (2000), which showed no evidence supporting the claim that extensive reading improves the reading comprehension of low proficiency learners.

One reason that the younger learners appeared to benefit less from extensive reading might be due to the fact there were a small number of studies in which young learners participated. It is plausible that extensive reading has beneficial effects on younger learners, as developing reading skills is a long process that requires a great deal of time-on-task to acquire. One of the

fundamental insights from the L1 literacy research is that cultivating fluent and effective reading skills requires the cultivation of effective reading habits, and this takes a great deal of time. Researchers working in this area need to take into account both the duration of the reading, the quality of the reading experiences, and the quality of instruction. In Chapter 2, it was noted that extensive reading is believed to enhance learners' knowledge of the contextual usage of L2 lexis. If extensive reading is introduced early in a curriculum, students have more opportunities to meet L2 lexis in context. This approach is important, as authorities (Alderson, 2000; Nation, 2009) agree that one of the keys to successful L2 learning is to increase the amount of comprehensible L2 input.

In a similar vein, the guidelines for setting up an extensive reading program reviewed in Chapter 2 include reading as much as possible. This advice remains at the heart of extensive reading approaches to this day. Other guidelines include reading for pleasure, reading at a relatively fast speed, providing students with a wide range of books including comic books and magazines, and providing time for students to read inside and outside of class.

Another important finding concerning the analysis of the age groups is that the number of studies in which junior high school and high school students

participated was greater for the group contrasts analysis than those with pre-post contrasts. This result might have arisen from the fact that junior high schools and high schools are part of the compulsory education system and for this reason, textbooks and syllabuses are set and difficult to change. Including a control group in these contexts can be difficult or impossible because each class is supposed to receive the same educational experiences. The differences between studies conducted in secondary schools and universities was also reflected in the time students spent engaging in extensive reading in class. For instance, Fujita and Noro (2009) had their high school participants read for 10 minutes in class; Matsui and Noro (2010) did the same with junior high school students. In contrast, Nakanishi and Ueda (2011) had their university students spend an entire 90-minute class engaged in extensive reading. University instructors in Japan generally have far more freedom to select class activities than secondary school teachers.

### **Length of Instruction**

The third research question asked whether length of instruction influenced test scores, and if so, what constituted an appropriate period of time for engaging in extensive reading. The results of a moderator variable analysis

indicated that one year of instruction for the group contrasts produced a small effect size ( $d = .52$ ), and the pre-post contrasts produced a medium effect ( $d = .74$ ). One semester of instruction for the group contrasts yielded a small effect ( $d = .36$ ), while the confidence interval for the effect size for the pre-post contrasts barely included zero (-0.01, 1.79). The effect sizes for two semesters of instruction included the smallest number of studies ( $n = 4, 3$  respectively), and thus produced negligible effects. Based on the results, longer instructional periods have the potential to improve students' reading proficiency, although it should be noted here again that the confidence intervals for all groups overlapped substantially, indicating no statistically significant difference between the three categories of instructional length (one semester vs. two semesters vs. one year). While the results in this study do not consistently support the use of extensive reading, there is agreement in the second language reading literature that reading over long periods of time is important. For instance, Grabe (2009) stated that "reading extensively, when done consistently over a long period of time, leads to better reading comprehension as well as improved abilities in several other language areas" (p. 328). More studies incorporating various instructional periods are required to confidently assess the

effectiveness of these periods. In addition, because second language learners often engage in extensive reading outside the classroom, it is crucial that accurate records are kept regarding the amount of reading completed, the order of the texts read, and the level of those texts read.

### **Outcome Measures**

The fourth research question concerned the kinds of tests that were utilized to determine the effects of extensive reading. As shown in Appendices 1, 2, and 3, summaries of past extensive reading studies dealt with sample sizes and gains, and the measures used in each study to determine gains were not discussed. Identifying the measurements utilized in past studies helps future researchers wishing to use the same measurements to replicate a study or compare their results with those obtained in other contexts. Researchers need to carefully choose or create valid and reliable tests in order to produce trustworthy results.

A wide range of tests was administered in the extensive reading studies included in this meta-analysis. Out of 34 studies, 20 different tests were administered to investigate the effect of extensive reading (see Table 14). The

most frequently used test was the EPER Placement/Progress Test ( $n = 5$ ) followed by Nation's Vocabulary Levels Test ( $n = 4$ ), and the TOEFL ( $n = 3$ ). The EPER placement/progress test, which has eight difficulty levels starting from version A, is a cloze test designed to measure learners' general English reading proficiency. It might have been used frequently because it was designed to assess learners' general English reading proficiency based on graded readers that learners read in extensive reading programs; furthermore, the name of the test itself, which includes the phrase *extensive reading*, suggests that it might be more effective than tests designed for other purposes (see Day & Bamford [1998] for a detailed description of the EPER tests).

These results clearly demonstrate that there has been a lack of replication studies in the extensive reading literature and they hint at the difficulty of comparing the results of studies that use different outcome measures. The analysis also shows that extensive reading researchers need to pay closer attention to the quality of the instruments they use. For instance, some researchers (e.g., Fujimori, 2006) used tests that had relatively few items, which raises issues of validity and reliability. One consequence of using

unreliable tests is that they obscure possible group differences and thereby reduce the magnitude of the effect sizes in studies such as this one.

The tests mentioned above were administered to assess three main areas: reading speed, reading comprehension, and vocabulary. Tables 12 and 13 show that the most frequently used test was a reading comprehension test. Studies using these tests produced medium effects ( $d = .63$  for group contrasts and  $d = .72$  for pre-post contrasts), while studies of reading speed produced a large effect of  $d = .98$  for group contrasts, but a non-significant effect for pre-post contrasts. Studies of vocabulary acquisition yielded a large effect ( $d = 1.25$ ) for pre-post contrasts; however, only five studies with a total sample size of 113 participants were included in this analysis; thus, more studies are needed in this area to confidently conclude that the effect size is large.

### **Deficiencies in Extensive Reading Research**

The last research question concerned the deficiencies in extensive reading research. Out of the 43 studies coded in this study, the number of books students read during the instruction was stated in only 18 studies with the mean number of books read ranging from 2 to 30. It is crucial that extensive reading

researchers clearly report the amount of reading each learner has completed given the important of the amount reading on acquisition. However, reporting the number of books and pages read is an extremely inaccurate approach to this issue give the great differences in the number of words in different books, and the differences in the number of words on one page of text in different graded readers. At a minimum, researchers should report the number of running words read, but an even better solution is to report the number of standard words read, as in Beglar, Hunt, and Kite's (2012) study.

In terms of research design, control groups were not included in approximately 40% ( $n = 17$ ) of the studies. According to Sanders (2010), "control is a standard of comparison for checking or verifying the results of an experiment or it can refer to an individual or group used for the purpose of comparison in an experiment" (p. 280). McDonough and McDonough (1997), in their discussion of experimental design, stated that "the greater the influence of such design measures, the nearer the research is to a true experiment" (p. 160). The advantages of experimental research can only be realized when proper designs are used and one aspect of rigorous designs is often the presence of one or more control groups.

In sum, extensive reading researchers need to include more detailed descriptive statistics, use control groups in their designs, use longitudinal designs, and conduct studies with elementary school age participants. Studies conducted in Europe are also needed given the current imbalance in the literature caused by the dominance of research conducted in Asia.

### **Comparison with Other Meta-Analyses of Extensive Reading**

The results of this study are discussed by comparing them with Wang (2010), which is the most comprehensive meta-analysis of extensive reading conducted prior to this study. Wang focused on moderator variables rather than the overall effect of extensive reading. Wang collected 35 studies published from 1977 to 2007 and divided them using six moderator variables: reading speed, vocabulary acquisition, reading comprehension, writing fluency, accuracy, and language proficiency. Analyses were conducted separately for controlled and uncontrolled (i.e., pre-post) experiments. The effect sizes produced by the uncontrolled experiments were larger than the ones produced by the controlled experiments. Large effects were obtained for reading speed, reading comprehension, writing fluency, accuracy, and language proficiency for

the uncontrolled experiments, while a large effect was found for reading speed and a medium effect was found for reading comprehension for the controlled experiments. College learners benefited significantly more than young learners where accuracy was concerned. In sum, Wang concluded that even though there are uncertainties over uncontrolled experiments, extensive reading has a positive effect on language learning and it should be a part of any foreign language curriculum.

This meta-analysis and Wang (2010) produced some similar findings. First, the effect sizes produced by uncontrolled experiments (pre-post contrasts) were larger than the ones by controlled experiments (group contrasts), which suggests that using control groups influences the robustness of the reported outcomes. Second, both studies found differences between adult learners and young learners, though those findings were not statistically significant in this study. This finding indicates that adult learners benefit more from extensive reading, but as Wang stated uncertainty about this conclusion remains because of the small number of extensive reading studies that have been conducted with young participants. Finally, the results for many of the outcome variables, such as reading speed and vocabulary included zero in the confidence intervals,

which means that trends rather than significantly significant differences are often found in studies of extensive reading.

The pedagogical significance of the meta-analyses conducted to date is that extensive reading does appear to have multiple benefits, but few of them are strong. It is still too early to make strong statements about the effects associated with moderator variables; however, extensive reading can clearly provide students with large amounts of input that they would not have access to if they only engaged in intensive reading. This is particularly true for students in EFL contexts, as they generally have few opportunities to encounter the target language outside the foreign language classroom. Another significant advantage of extensive reading is its flexibility. Students can read virtually anywhere and at any time. Once they develop the habit of reading, they can engage in what Nuttall (2005) called the “virtuous circle of the good reader” (p. 127).

## CHAPTER 7

### CONCLUSION

#### Summary of the Findings

The overall effect sizes for extensive reading treatments in this meta-analysis were small for group contrasts ( $d = .46$ ) and medium for pre-post contrasts ( $d = .71$ ). These findings indicate that extensive reading is an effective instructional approach for improving reading proficiency. Most extensive reading research has been conducted with university students, and most studies were conducted for less than six months; therefore there is a need for more studies in which the participants are in elementary or secondary school, and for longitudinal studies conducted for one or more years. Various tests were used to assess students' improvement, but this made it difficult to compare the results among studies. Most of the studies reviewed in this meta-analysis were conducted in Asia; therefore, there is a need for more studies to be conducted in regions such as Europe and North America so that the effects of extensive reading in various educational contexts can be better understood.

## **Limitations**

There are three main limitations of this study. First, this synthesis included a small number of studies compared to other meta-analyses (e.g., Masgoret & Gardner, 2003; Norris & Ortega, 2000; Taylor, Stevens, & Asher, 2006) and because the data were divided into group contrasts and pre-post contrasts, the sample size became even smaller. The small number of studies also affected the moderator variable analysis. Eighteen studies with group contrasts and 16 studies with pre-post contrasts were divided into subgroups; some of the effect sizes produced wide confidence intervals, some of which included zero. The confidence intervals for all groups overlapped substantially with each other (e.g., one semester vs. two semesters vs. one year). This made it difficult to interpret the results with confidence. In the field of SLA, it is common to conduct research using intact classes where there are approximately 30 students in class. For this reason, most studies have less than 100 participants. This small number of participants reduces statistical power and leads to smaller effect sizes.

A second limitation is that no studies with non-significant findings were included in this synthesis. This might lead to publication bias and might be the

reason for the lack of significance found in the two homogeneity tests. This is a difficult problem to overcome, as it is unlikely that refereed journals will begin to publish studies with non-significant findings in the near future

The third limitation is that most of the studies were conducted in Asia. The effect of extensive reading might be considerably greater with participants with first languages genetically related to English (e.g., German or Dutch), as their L1s would share many similarities with English (e.g., orthography, lexis, syntax). More studies conducted in non-Asian contexts would provide a more well rounded view of the efficacy of extensive reading and shed light on the influence of learners' L1s as well.

### **Implications for Research Methodology**

In order to enable future researchers who conduct meta-analyses to provide more accurate results, extensive reading researchers should consider the following seven recommendations.

1. Researchers should include detailed descriptive statistics when reporting their results. The major reason for the small sample size in this meta-analysis is

because many primary studies lack the information needed to calculate effect sizes. Most authors who have conducted experimental or quasi-experimental studies include the mean scores for each group, but they generally fail to provide the standard deviations. This problem applies to pre-post designs as well. For instance, the study by Sheu (2003) was excluded from the meta-analysis due to the lack of necessary information to calculate effect sizes.

Dinsmore (2006) also stated that “future quantitative studies should provide not only the means and standard deviations, but also group size(s), sum of squares, *F* and *t* statistic values, and exact *p* value of significant and non-significant results, so that they can be included in future meta-analysis” (p. 80).

Researchers should also report effect sizes. APA (2010) recommends including a measure of effect size when reporting results because a *p* value alone does not connote an important finding. If effect sizes were reported more frequently, studies with nonsignificant findings might be published because even nonsignificant findings can be accompanied by reasonably large effect sizes.

2. There is a need for replication studies because such studies enable researchers to more precisely identify the effects of an intervention and the stability of those

effects across different educational contexts. As the number of replications increases, the results can be considered more reliable. This problem is partly in the hands of journal editors and reviewers, as they are the gatekeepers who must accept well conducted replication studies.

3. Researchers should include one or more control groups in their studies.

Sixteen studies included in the pre-post contrasts analysis did not have a control group. When investigating the effects of a treatment, researchers need to have something to compare to and that is where control groups play a crucial role.

Nine studies in this meta-analysis included two or more control groups, which indicates that more researchers are aware of the importance of having multiple control groups, especially in longitudinal studies where attrition is a danger.

4. More longitudinal studies are needed. The difficulty of conducting longitudinal research is indicated by the small number of studies ( $n = 16$ ) conducted from six months to one year. Only one study (Mason, 2003), a doctoral dissertation, was conducted for more than one year. Although many researchers (e.g., Horst, 2005; Iwahori, 2008; Kweon & Kim, 2008; Lai, 1993)

conducted research over about one semester, I would suggest that most extensive reading studies should be conducted over a period of at least one academic year. Through my experience of teaching extensive reading classes, students initially struggle to read in the L2, and they gradually begin to increase their reading pace. They then start to enjoy reading and eventually start to read more; it is at this point that the effects of extensive reading can appear more clearly.

5. Studies with elementary school age participants are needed. The lack of extensive reading studies with young participants was brought to light in this meta-analysis given that university students made up 56% of the participants in the extensive reading studies included in this meta-analysis. English education is now a part of the elementary school curriculum in countries such as Korea and Japan, so the time is ripe for learners in this age group to be studied. If the joy of reading is developed in learners at an early age, they will presumably continue to read books regularly and thereby reap greater benefits as the years pass.

6. Studies with large sample sizes need to be conducted. Figures 1 and 2 indicated clearly that most studies included in the meta-analysis were small-sample studies. One challenge of increasing the size of research samples is the requirement to provide a larger number of books as the number of participant increases. One way to increase sample size is to have multiple experimental groups and control groups. For example, Beglar et al. (2012) had multiple experimental groups and Nakanishi and Ueda (2011) utilized two experimental groups and two control groups. This approach is preferable because it guards against attrition in longitudinal studies.

7. Research need to provide more accurate counts of how much learners read. Out of the 43 studies coded in this study, only 18 listed the number of books students read and nine reported the number of words students read. While counts of books and pages can be reported, they are imprecise measures because the number of words can differ considerably from book to book and from page to page in books written at different levels. Therefore, in addition to reporting how many books participants read, researchers should also report the number of running words read or number of standard words read.

### **Suggestions for Future Research**

This study has shown that the field has witnessed an increase in the number of studies published over the past 20 years, and all signs suggest that this trend will continue. The following aspects of extensive reading can be investigated in the future.

1. Researchers can investigate the degree to which ER helps learners build connections between lexical units such as words that form collocations and lexical bundles.
2. Research investigating different approaches to extensive reading, such as  $i + 1$  level vs.  $i - 1$  level, or the effects of re-reading books is needed. Different types of extensive reading might differentially affect the acquisition of different aspects of the target language.
3. More studies of motivation and other affective issues, such as reading self-efficacy, and their relationship to extensive reading need to be conducted.

Waring (2001), who argued that these areas are underresearched, stated that

“Research into language gains and gains in affect (e.g., confidence and motivation) from ER in second languages is still in its infancy, is quite fragmented, and is rather difficult to interpret when looking for concrete ‘evidence’” (p. 4). Some research (e.g., Nishino, 2007; Takase, 2004), however, provides a good foundation for investigating these aspects of extensive reading.

4. Although previous research (e.g., Norris & Ortega, 2000; Spada & Tomita, 2010) has indicated that explicit instruction is more effective than implicit instruction, extensive reading is a fundamentally implicit instructional method. Researchers could, however, combine explicit instruction with extensive reading to further enhance the effects of extensive reading on areas such as spelling, lexical acquisition, and genre organization. For instance, Nakanishi and Ueda (2011) combined extensive reading and shadowing to investigate the combined effect.

## **Final Conclusions**

A major benefit of a meta-analysis is that it can provide far more reliable and informative results than any single study. Meta-analyses can also show the tendency of past research and suggest future research directions; moreover, researchers can use the results obtained not only as statistical outcomes, but also as a means to create alternative hypotheses for future research.

The available extensive reading research to date suggests that extensive reading does improve students' reading proficiency. Hopefully, the issues discussed in this study have offered useful insights for extensive reading researchers.

## REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Abelson, R. P. (1997). A retrospective on the significance test ban of 1999 (if there were no significant tests, they would be invented). In L. L. Harlow, S. A. Mulaik & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 117-141). Mahwah, NJ: Erlbaum.

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning, 21*, 199-226.

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strength and weaknesses. *Review of Educational Research, 82*(4), 436-476.

\*Al-Homoud, F., & 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), 383-401.

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

American Psychological Association. (2010). *Publication manual of the American psychological association* (6th ed.). Washington, DC: American Psychological Association.

Anderson, N. J. (2009). ACTIVE Reading: The research base for a pedagogical approach in the reading classroom. In Han, Z., & Anderson, N. J., *Second language reading research and instruction: Crossing the boundaries*. (pp. 117-143). Ann Arbor, MI: The University of Michigan Press.

- Bamford, J., & Day, R. R. (2004). *Extensive reading activities for teaching language*. Cambridge, England: Cambridge University Press.
- Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58.
- Becker, B. J. (1994). Combining significance levels. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 215-230). New York, NY: Russell Sage Foundation.
- \*Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese EFL learners' reading rates. *Language Learning, 62*(3), 665-703.
- \*Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix, 1*(1); Retrieved from <http://www.readingmatrix.com/articles/bell/index.html>
- Bernhardt, E. B., & 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.
- Blok, H. (1999). Reading to young children in educational settings: A meta-analysis of recent research. *Language Learning, 49*(2), 343-371.
- Boning, R. A. (1977). *Multiple skills series midway placement test*. Baldwin, New York: Barnell and Loft.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley & Sons.
- Burrows, L. (2012). *The effects of extensive reading and reading strategies on reading self-efficacy*. Unpublished doctoral dissertation, Temple University, Tokyo.

- Camiciottoli, B. C. (2001). Extensive reading in English: Habits and attitudes of a group of Italian university EFL students. *Journal of Research in Reading*, 24(2), 135-153.
- Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics*, 12(2), 159-179.
- Carver, R. (1987a). *User's guide for the accuracy level test*. Kansas City, MO: Revrac.
- Carver, R. (1987b). *User's guide for the rate level test*. Kansas City, MO: Revrac.
- \*Cha, J. (2009). The effects of extensive reading on enhancing vocational high school students' L2 vocabulary & reading rates. *English Teaching*, 64(3), 3-30.
- Cho, K.-S., & Krashen, S. D. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, 37(8), 662-667.
- \*Cirocki, A. (2009). Implementing the ER approach to literature in the EFL secondary school classroom: An action research study. In A. Girocki (Ed.), *Extensive reading in English language teaching* (pp. 521-545). Munich, Germany: Lincom.
- Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 225-237). Cambridge, England: Cambridge University Press.
- Cobb, T. (2010). Web Vocabprofile (Version 2.5) [Computer software and manual], an adaptation of Nation & Heatley's (1994) Range.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

- Comprehensive Meta-Analysis (2007). (Version 2.2) [Computer software]. Englewood, NJ: Biostat.
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.
- Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York, NY: Russell Sage Foundation.
- Cornell, J., & Murlow, C. (1999). Meta-analysis. In A. Herman & M. Gideon (Eds.), *Research mythology in the social behavioral, and life sciences* (pp. 285-323). London, England: Sage.
- Davis, C. (1995). Extensive reading: An expensive extravagance? *ELT Journal*, 49(4), 329-336.
- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge, England: Cambridge University Press.
- Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, 14, 136-141.
- \*de Morgado, N. F. (2009). Extensive reading: Students' performance and perception. *The Reading Matrix*, 9(1), 31-43.
- Dinsmore, T. H. (2006). Principles, parameters, and SLA: A retrospective meta-analytic investigation into adult L2 learners' access to Universal Grammar. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 53-90). Amsterdam, Holland: Benjamins.
- Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. Doughty & M. Long (Eds.). *The handbook of second language acquisition* (pp. 256-310). Oxford, England: Blackwell.

- Edinburgh Project on Extensive Reading (EPER), not dated. EPER publications: Supporting English Reading Schemes for Schools and Colleges World-Wide [Brochure], University of Edinburgh, Edinburgh.
- Elley, W. B. (1986). Evaluating the benefits of book floods in second language learning. In V. Bickley (Ed.), *Future directions in English language teachers education: Asia and Pacific perspectives* (pp. 211-225). Hong Kong, China: Education Department.
- Elley, W. B., & Mangubhai, F. (1981). *The impact of a book flood in Fiji primary schools*. Wellington, New Zealand: NZCER.
- Elley, W. B., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly*, 19, 53-67.
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. New York, NY: Cambridge University Press.
- Fields, A. P. (2003). Can meta-analysis be trusted? *Psychologist*, 16, 642-645.
- \*Fujimori, C. (2006). The effects of an extensive reading program on reading and listening comprehension among senior high school students. *KATE Journal*, 20, 13-23.
- \*Fujita, K., & Noro, T. (2009). The effects of 10-minute extensive reading on the reading speed, comprehension and motivation of Japanese high school EFL learners. *Annual Review of English Language Education in Japan*, 20, 21-30.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher*, 5, 3-8.
- Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, 51(1), 1-50.

- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge, England: Cambridge University Press.
- Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd ed.). Harlow, England: Pearson.
- \*Greenberg, D., Rodrigo, V., Berry, A., Brinck, T., & Joseph, H. (2006). Implementation of an extensive reading program with adult learners. *Adult Basic Education, 16*(2), 81-97.
- Greenhouse, J. B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 383-398). New York, NY: Russell Sage Foundation.
- Hafiz, F. M., & Tudor, I. (1989). Extensive reading and the development of language skills. *ELT Journal, 43*(1), 4-13.
- Hafiz, F. M., & Tudor, I. (1990). Graded readers as an input medium in L2 learning. *System, 18*, 31-42.
- Han, Z., & Anderson, N. J. (2009). *Second language reading research and instruction: Crossing the boundaries*. Ann Arbor, MI: The University of Michigan Press.
- \*Hayashi, K. (1999). Reading strategies and extensive reading in EFL classes. *RELC Journal, 30*(2), 114-132.
- Hedges, L. (1994). Statistical consideration. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 29-38). New York, NY: Russell Sage Foundation.
- \*Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review, 61*(3), 355-382.
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York, NY: Russell Sage Foundation.

- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, 26, 219-244.
- In'nami, Y., & Koizumi, R. (2010). Database selection guidelines for meta-analysis in applied linguistics. *TESOL Quarterly*, 44, 169-184.
- \*Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language*, 20, 70-91.
- Janopoulos, M. (1986). The relationship of pleasure reading and second language writing proficiency. *TESOL Quarterly*, 20(4), 763-768.
- Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165-211). Amsterdam, Holland: Benjamins.
- Keck, C. M., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition: A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91-131). Amsterdam, Holland: Benjamins.
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, 28(1), 77-96.
- Kim, J.-R. (2011). *Meta-analysis of classroom-based extensive reading programs*. Paper presented at the First Extensive Reading World Congress, Kyoto, Japan.
- Koda, K. (2004). *Insights into second language reading*. New York, NY: Cambridge University Press.

- Krashen, S. (2007). Extensive reading in English as foreign language by adolescents and young adults: A meta-analysis. *The International Journal of Foreign Language Teaching*, 3, 23-29
- Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95, 3-12.
- \*Kweon, S.-O., & Kim, H.-R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language*, 20(2), 191-215.
- \*Lai, F.-K. (1993). The effect of a summer reading course on reading and writing skills. *System*, 21, 87-100.
- \*Lao, C. Y., & Krashen, S. (2000). The impact of popular literature study on literacy development in EFL: More evidence for the power of reading. *System*, 28, 261-270.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5, 223-250.
- Laufer, B., & Nation, I. S. P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, 16, 33-51.
- Lee, S.-K., & Huang, H.-T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies of Second Language Acquisition*, 30, 307-331.
- \*Lee, S.-Y. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal*, 38(2), 150-170.
- Leung, C. Y. (2002). Extensive reading and language learning: A diary study of a beginning learner of Japanese. *Reading in a Foreign Language*, 14, 66-81.

- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309-365.
- \*Lin, L-F. (2010). Senior high school students' reading comprehension of graded readers. *Journal of Language Teaching and Research, 1*(1), 20-28.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- \*Lituanas, P, M., Jacobs, G. M., & 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.
- Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading, 32*, 199-214.
- Lynch, J. (2002). Parent's self-efficacy beliefs, parents' gender, children's reader self-perceptions, reading achievement and gender. *Journal of Research in Reading, 25*, 54-67.
- Lyster, R., & Saito, Y. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies of Second Language Acquisition, 32*, 265-302.
- Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning, 53*(1), 123-163.
- \*Mason, B. (2003). *A study of extensive reading and the development of grammatical accuracy by Japanese university students learning English*. Unpublished doctoral dissertation, Temple University, Tokyo.

- \*Mason, B. (2007). The efficacy of self-selected reading and hearing stories on adult second language acquisition. *The Proceedings of the Sixteenth International Symposium on English Teaching* (pp. 630-633). Taipei: Crane Publishing Company. Retrieved Aug 1, 2012, from <http://www.benikomason.net/articles/eta2007/all.html>
- \*Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, 25(1), 91-102.
- Masuhara, H., Kimura, T., Fukada, A., & Takeuchi, M. (1996). Strategy training or/and extensive reading? In T. Hickey & J. Williams (Eds.), *Language, education, and society in a changing world* (pp. 263-274). Clevedon, UK: Multilingual Matters.
- \*Matsui, T., & Noro, T. (2010). The effects of 10-minute sustained silent reading on the junior high school EFL learners' reading fluency and motivation. *Annual Review of English Language Education in Japan*, 21, 71-80.
- McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, 71(1), 173-180.
- McDonough, J., & McDonough, S. (1997). *Research methods for English language teachers*. London: Arnold.
- Mikulecky, B., & Jeffries, L. (2005). *Reading power* (3rd ed.). New York, NY: Addison-Wesley Longman.
- Mori, S. (1999). The role of motivation in the amount of reading. *Temple University Japan Working Papers in Applied Linguistics*, 14, 51-68.
- Mori, S. (2002). Redefining motivation to read in a foreign language. *Reading in a Foreign Language*, 14, 91-110.
- Murphy, K. R., & Myers, B. (2004). *Statistical power analysis*. Mahwah, NJ: Erlbaum.

- Muto, N. W. (2006). Why extensive reading? *Journal of School of Foreign Languages, Nagoya University of Foreign Studies*, 30, 1-16.
- Nakanishi, T. (2005). A book in question! Intertwining receptive and productive skills using graded readers. *The Language Teacher*, 29(6), 35-36.
- Nakanishi, T. (2008). Learner perceptions and attitudes toward learning styles in Japan and Taiwan. *JLTA (Japan Language Testing Association) Journal*, 11, 21-40.
- \*Nakanishi, T., & Ueda, A. (2011). Extensive reading and the effect of shadowing. *Reading in a Foreign Language*, 23(1), 1-16.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York, NY: Newbury House.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. New York, NY: Routledge.
- Nation, I. S. P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, 12(2), 355-380.
- National Institute of Child Health and Human Development (2000). Report of the National Reading Panel. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: U.S. Government Printing Office.
- Nishino, T. (2007). Beginning to read extensively: A case study with Mako and Fumi. *Reading in a Foreign Language*, 19, 76-105.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.

- Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, 51(1), 157-213.
- Norris, J. M., & Ortega, L. (Eds.). (2006a). *Synthesizing research on language learning and teaching*. Amsterdam, Holland: Benjamins.
- Norris, J. M., & Ortega, L. (Eds.). (2006b). The value and practice of research synthesis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3-50). Amsterdam, Holland: Benjamins.
- Norris, J. M., & Ortega, L. (2007). The future of research synthesis in applied linguistics: Beyond art or science. *TESOL Quarterly*, 41, 805-815.
- Nuttall, C. (2005). *Teaching reading skills in a foreign language*. Oxford, England: Macmillan Education.
- Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, 30, 85-110.
- Patall, E., Cooper, H., & Robinson, J. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, 134, 270-300.
- Perkins, K., Brutton, S. R., & Pohlmann, J. T. (1989). First and second language reading comprehension. *RELC Journal*, 20(2), 1-9.
- Pigott, T. D. (2012). *Advances in meta-analysis*. New York, NY: Springer.
- Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of Clockwork Orange study using second language acquirers. *Reading in a Foreign Language*, 5(2), 271-275.

- Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61, 993-1038.
- Plonsky, L. (2012, June 18). Meta-analysis in Applied Linguistics. Retrieved from [http://oak.ucc.nau.edu/ldp3/bib\\_metaanalysis.html](http://oak.ucc.nau.edu/ldp3/bib_metaanalysis.html)
- Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcome: The case of interaction research. *Language Learning*, 61(2), 325-366.
- Plonsky, L., & Oswald, F. L. (2012). How to do a meta-analysis. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 275-295). London, England: Blackwell.
- Prowse, P. (2002). Top ten principles for teaching extensive reading: A response. *Reading in a Foreign Language*, 14(2), 142-145.
- Raudenbush, S. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-322). New York, NY: Russell Sage Foundation.
- \*Rezaee, A. A., & Nourzadeh, S. (2011). Does extensive reading improve EFL learners' processing ability? *Theory and Practice in Language Studies*, 1(9), 1167-1175.
- \*Robb, T. N., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, 5, 239-251.
- \*Rodrigo, V., Krashen, S., & Gribbons, B. (2004). The effectiveness of two comprehensible-input approaches to foreign language instruction at the intermediate level. *System*, 32, 53-60.
- Rolstad, K., Mahoney, K., & Glass, G. (2005). Weighing the evidence: A meta-analysis of bilingual education in Arizona. *Bilingual Research Journal*, 29, 43-67.

- Rosenthal, R. (1979). The 'file drawer' problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R. (1987). *Judgment studies: Design, analysis, and meta-analysis*. New York, NY: Cambridge University Press.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York, NY: Russell Sage Foundation.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118(2), 183-192.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1-20.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). (Eds.). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, England: John Wiley & Sons.
- Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133-164). Amsterdam, Holland: Benjamins.
- Sahari, M. (1997). Elaboration as a text-processing strategy: a meta-analytic review. *RELC Journal*, 28(1), 15-27.
- Sakai, K., & Kanda, M. (2005). *Kyoushitudeyomu eigo 100mango* [Reading one million words in the classroom]. Tokyo, Japan: Taishukan Shoten.
- Sanders, L. D. (2010). *Discovering research methods in psychology: A student's guide*. London, England: BPS Blackwell.

- Schmitt, N. (2008). Instructed second language vocabulary. *Language Teaching Research*, 12(3), 329-363.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, 18, 55-88.
- Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261-281). New York, NY: Russell Sage Foundation.
- Sheu, S. P.-H. (2003). Extensive reading with EFL learners at beginning level. *TESL Reporter*, 36, 8-26.
- Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning*, 63, 296-329.
- \*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*, 2(2), 12-15. Retrieved Aug 11, 2007, from <http://www.tprstories.com/ijflt/IJFLTFall06.pdf>
- Spada, N., & Tomita, Y. (2010). Interactions between types of instruction and types of language feature: A meta-analysis. *Language Learning*, 60(2) 263-308.
- \*Taguchi, E., Takayasu-Maass, M., & Gorsuch, J. G. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language*, 16, (2), 70-96.
- \*Takase, A. (2004). *The effects of extensive reading on the motivation of Japanese high school students*. Unpublished doctoral dissertation, Temple University, Tokyo.

- \*Takase, A. (2007a). Japanese high school students' motivation for extensive L2 reading. *Reading in a Foreign Language*, 19, 1-18.
- Takase, A. (2007b). Extensive reading in the Japanese high school setting. *The Language Teacher*, 31(5), 7-10.
- \*Takase, A. (2009). The effects of SSR on learners' reading attitudes, motivation and achievement: A quantitative study. In A. Cirocki (Ed.), *Extensive reading in English language teaching* (pp. 547-559). Munich, Germany: Lincom.
- Takase, A. (2010). *Eigo tadoku tacyou shidou manyuaru* [English extensive reading and listening teaching manual]. Tokyo, Japan: Taisyukan.
- \*Tanaka, H., & 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. Retrieved Oct 27, 2008  
from [http://www.readingmatrix.com/articles/tanaka\\_stapleton/article.pdf](http://www.readingmatrix.com/articles/tanaka_stapleton/article.pdf)
- Taylor, A., Stevens J., & Asher, W. (2006). The effects of explicit reading strategy training on L2 reading comprehension: A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 213-244). Amsterdam, Holland: Benjamins.
- \*Tsang, W.-K. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics*, 17, 210-233.
- Valentine, J. C., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.
- Wang, L.-J. (2010). *A meta-analysis of empirical studies on the effects of extensive reading*. Unpublished master's thesis, National Tsing Hua University, Taiwan.

- Waring, R. (2001). Research in extensive reading. *Notre Dame Seishin University: Studies in Foreign Languages and Literature*, 25(1), 44-67
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 130-163.
- White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 41-55). New York, NY: Russell Sage Foundation.
- Williams, E. (2007). Extensive reading in Malawi: Inadequate implementation or inappropriate innovation? *Journal of Research in Reading*, 30(1), 59-79.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3, 215-229.
- \*Yamamoto, Y. (2011). Bridging the gap between receptive and productive vocabulary size through extensive reading. *The Reading Matrix*, 11(3), 226-242.
- Yamashita, J. (2002). Mutual comprehension between L1 reading and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading*, 25(1), 80-94.
- Yamashita, J. (2004). Reading attitudes in L1 and L2, and their influence on L2 extensive reading. *Reading in a Foreign Language*, 16, 1-19.
- Yamashita, J. (2007). The relationship of reading attitudes between L1 and L2: An investigation of adult EFL learners in Japan. *TESOL Quarterly*, 41, 81-105.
- \*Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, 36, 661-672.

\*Yang, A. (2001). Reading and the non-academic learner: A mystery solved. *System*, 26, 451-466.

\*Zimmerman, C. B. (1997). Do reading and interactive vocabulary instruction make a difference?: An empirical study. *TESOL Quarterly*, 31(1), 121-140.

## **APPENDICES**

**APPENDIX A**  
**SUMMARY OF EXTENSIVE READING STUDIES**  
**(Day & Bamford, 1998, p. 34)**

Report	Population	Results
Elley & Mangubhai (1981)	EFL; primary; Fiji	Gains in reading and general proficiency, including listening and writing; growth in positive affect.
Janopoulos (1986)	ESL; university; USA	Gains in writing proficiency.
Hafiz & Tudor (1989);	ESL; adolescents;	Gains in reading proficiency, positive affect, and general linguistic competence, including writing; slight, nonsignificant increase in vocabulary.
Tudor & Hafiz (1989)	England	
Pitts et al. (1989)	ESL; adults; USA	Gains in vocabulary
Robb & Susser (1989)	EFL; university; Japan	Gains in reading proficiency and positive affect.
Hafiz & Tudor (1990)	EFL; university; Pakistan	Gains in vocabulary and writing.
Elley (1991)	EFL; primary; Singapore	Gains in reading proficiency and positive affect.
Lai (1993a, 1993b)	EFL; secondary; Hong Kong	Gains in reading proficiency and vocabulary.
Cho & Krashen (1994)	ESL; adults; USA	Gains in reading proficiency, vocabulary, positive affect, and oral skills.
Rodrigo (1995)	Spanish; university; USA	Gains in positive affect; no statistically significant gains in vocabulary.
Mason & Krashen (1997)	EFL; university; Japan	Gains in reading proficiency, positive affect, and writing.

**APPENDIX B**  
**SUMMARY OF EXTENSIVE READING STUDIES (Horst, 2005, p. 359)**

Study	Context & population	Reading materials	Mean amount read per week	Other English input	Results: change in ...
Robb & Susser (1989)	125 EFL Japanese university	simplified for NSs	18 pages	EFL classes	<ul style="list-style-type: none"> <li>• reading comprehension</li> <li>• speed</li> <li>• vocabulary</li> </ul>
Hafiz & Tudor (1989)	16 ESL Pakistani secondary school students	graded ESL readers	4.5 hours	UK milieu	<ul style="list-style-type: none"> <li>• reading comprehension</li> <li>• writing</li> </ul>
Hafiz & Tudor (1990)	25 EFL Pakistani secondary school students	graded ESL readers	4 hours	EFL classes	<ul style="list-style-type: none"> <li>• essay writing</li> </ul>
Lai (1993)	250+ EFL Hong Kong secondary school students	graded ESL readers	4-5 books	English in milieu?	<ul style="list-style-type: none"> <li>• reading comprehension</li> <li>• speed</li> <li>• essay writing</li> <li>• vocabulary</li> </ul>
Cho & Krashen (1994)	4 ESL adults	Simplified for NSs	2-6 books	US milieu	<ul style="list-style-type: none"> <li>• essay writing</li> <li>• vocabulary</li> </ul>
Tsang (1996)	48 EFL Hong Kong secondary school students	graded ESL readers	3 books	EFL classes	<ul style="list-style-type: none"> <li>• essay writing</li> </ul>
Mason & Krashen (1997)	20, 128, 76 EFL Japanese university (3 experiments)	graded ESL readers (-t-unsimplified in exp. 2)	1-2 books in experiment 1	EFL classes in experiments 2 & 3	<ul style="list-style-type: none"> <li>• cloze</li> <li>• reading comprehension</li> <li>• summary writing</li> <li>• speed*</li> <li>• cloze**</li> </ul>
Walker (1997)	43 ESL adults	graded ESL readers	1 book	UK milieu	<ul style="list-style-type: none"> <li>• cloze**</li> </ul>
Lituaras, Jacobs, & Renandya (2001)	30 ESL Filipino secondary school students	various unsimplified ? materials	1.5 hours	ESL classes & English in milieu	<ul style="list-style-type: none"> <li>• reading comprehension</li> <li>• oral reading</li> </ul>
Renandya, Rajan, & Jacobs (1999)	49 EFL Vietnamese government employees	graded ESL readers	91 pages	EFL classes	<ul style="list-style-type: none"> <li>• integrated skills**</li> </ul>
Bell (2001)	14 EFL Yemeni government employees	graded ESL readers	1.5 + (?)hours	EFL classes	<ul style="list-style-type: none"> <li>• speed</li> <li>• cloze</li> <li>• reading comprehension</li> </ul>

**APPENDIX C**  
**SUMMARY OF EXTENSIVE READING STUDIES**  
**(Iwahori, 2008, p. 71)**

Study	<i>N</i>	Population	Results
<b>Vocabulary</b>			
Cho & Krashen, 1994	4	ESL; adults; USA	Gains
Horst, 2005	21	ESL; adults; Canada	Gains
Pitts, White, & Krashen, 1989	51	ESL; adults; USA	Gains
<b>Writing</b>			
Hafiz & Tudor, 1990	25	EFL; high school; Pakistan	Gains
Tsang, 1996	48	EFL; junior high school, high school; Hong Kong	Gains
<b>Reading comprehension</b>			
Masuhara, Kimura, Fukuda, & Takeuchi, 1996	46	EFL; college; Japan	Gains
<b>Reading comprehension and speed</b>			
Bell, 2001	14	EFL; young adults; Yemen	Gains
Robb & Susser, 1989	(About 62)	EFL; university; Japan	Gains
Sheu, 2003	65	EFL; junior high school; Taiwan	Gains
Taguchi, Takayasu-Maass, & Gorsuch, 2004	10	EFL; university; Japan	No gains
<b>Reading comprehension, writing, and speed</b>			
Lai, 1993	266 <sup>a</sup>	EFL; junior high school; Hong Kong	Gains
<b>Reading comprehension, writing, and attitude</b>			
Mason & Krashen, 1997 (3 studies)	20, 71, 76	EFL; university; Japan	Gains

*Note.* *N* = the number of participants in the ER treatment groups.

<sup>a</sup>The actual number of participants whose reading speed was analyzed was 207.