THE EFFECTS OF SECOND-LANGUAGE REPEATED READING ON READING

COMPREHENSION AND INCIDENTAL VOCABULARY ACQUISITION

by

Cheng-Ling Chen

Dissertation Committee:

Professor ZhaoHong Han, Sponsor
Professor Vivian Lindhardsen

Approved by the Committee on the Degree of Doctor of Education

Date _____17 October 2018_____

Submitted in partial fulfillment of the
requirements for the Degree of Doctor of Education in
Teachers College, Columbia University

2018

ABSTRACT

THE EFFECTS OF SECOND-LANGUAGE REPEATED READING ON READING

COMPREHENSION AND INCIDENTAL VOCABULARY ACQUISITION

Cheng-Ling Chen

Reading in a second language (L2) is considered a necessary skill in increasingly globalized societies. Not only is reading for purposes of comprehension necessary for survival, also reading in an L2 is an important means by which L2 acquisition occurs, particularly where vocabulary is concerned. Consequently, there is a strong demand for L2 research to investigate the instructional conditions that will best promote reading comprehension and vocabulary through efficient and effective reading strategies.

The current study addressed this dual need, reading comprehension and vocabulary acquisition, through an investigation of a particular type of pedagogical intervention, repeated reading (RR; i.e., multiple encounters with the same text), with high school English language learners in Taiwan. The study examined the effects of three conditions – Unassisted RR (repeated reading only), Assisted RR (repeated reading plus listening), and Control – on the participants' reading comprehension and incidental vocabulary acquisition through a pre-test, post-test, and delayed post-test design.

The results of the data from 42 participants suggested that L2 RR did not promote reading comprehension, nor did it contribute to a transfer of practice effect to new text in terms of reading comprehension. However, there were statistically significant incidental vocabulary gains and retention for the Unassisted RR group and some vocabulary gains for the Assisted RR group. When the percentage of unknown words of a text reached 10% and the participants were not provided with additional support, five repeated encounters with the text (over eight treatment sessions) were found to be inadequate in promoting reading comprehension. Nonetheless, the participants provided with such challenging condition still benefited from the incidental vocabulary acquisition. Findings may imply that a certain threshold of proficiency (e.g., percentage of known words of a text) is necessary for the beneficial effects of repeated reading to support comprehension. Additionally, RR alone may still be insufficient and additional support to RR may still be necessary for L2 learners dealing with difficult texts.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor and dissertation sponsor, Professor ZhaoHong Han, for her tremendous support, encouragement, and guidance throughout the process. She is an absolute inspiration and superwoman – dedicated, bright, creative, energetic, pioneer, and always passionate about knowledge, research, and teaching – and she constantly challenges herself and her students so that we can continue to grow intellectually. Second, I would like to thank my other committee members, Professors Vivian Lindhardsen, Bryan Keller, and Dolores Perin for their support and encouragement, and for providing numerous hours of assistance and valuable feedback on my dissertation. Additionally, I would like to thank the students and teachers at the cooperating school; this study would not have been possible without their assistance in support of EFL education and research.

I am extremely grateful to have the support from my SLA doctoral seminar fellows. My special thanks go to Ya-Yun Anny Sun and Farah Akbar for being great friends and for taking care of me, and special thanks to Timothy Hall for providing tremendous encouragement and hours of feedback on my dissertation. My sincere thanks also go to Shafinaz Ahmed, Philipp Choong, Kaylee Fernandez, Rosette Finneran, Jordan Gonzalez, Ji-Yung Jung, EunYoung Kang, Peter Kim, Adrienne Wai Man Lew, Mi Sun Park, Shaoyan Qi, Natalia Sáez, Sarah Sok, Haimei Sun, and Rainie Zhang. I enjoyed interacting with them, both intellectually and personally, and will really miss our SLA family and support group.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

1.1 Background

Reading is a unique human activity that sets us apart from other species. The influence and importance of reading to mankind cannot be overemphasized. It has changed our cognitive capacity, expanded our knowledge, and moreover, altered the avenue of individual and cultural development (Aebersold & Field, 1997). Approximately 80% of the world has the ability to read at a basic level, and a large percentage of these people can read at a high level of comprehension (Grabe, 2009). Reading is an important literacy skill to develop, because people are required to read and comprehend text for a variety of purposes. In order to meet the functional needs of daily life, many tasks require comprehension of texts such as newspapers and forms, emails or text messages, or product descriptions. In professional and educational contexts, reading plays an essential role for successful performance and provides the most common means of knowledge building. Such purposes for reading can be quite demanding and often require the ability to understand, interpret, integrate, synthesize, evaluate, and criticize the information from texts (Grabe, 2009). Furthermore, the advancements in technology and electronic communication have increased the large amount of information imposed upon individuals and therefore the demands for effective reading skills (Grabe, 2009). Hence, a high level of reading competency in both print and digital media is necessary for participation in 21st-century life (National Research Council, 2012). For the large

majority of individuals, reading is seen as an essential ability to further one's goals and better one's life (Grabe & Stoller, 2011). Without a strong reading competency, the chances for success and a quality life are limited (Grabe, 2009).

A strong reading competency is especially necessary due to the rise of globalization in today's society. The rise of globalization and the advances in transportation and technology have shortened the physical and virtual distance between people and places. Progress is seen through the emergence of global business markets, immigration to foreign countries, the popularity of traveling or studying abroad, and the development of the Internet (Chen, 2013). As a result, the usage of multiple languages occurs on a daily basis, making reading competency in another language increasingly important. According to statistics provided by the Linguistic Society of America, the majority of nations in today's society are multilingual, with an estimate of over five thousand languages used in 160 nations around the world (Valdes, 2012). Furthermore, people that are multilingual outnumber people that are monolingual (Tucker, 1998). In terms of the English language, approximately 400 million people are native speakers but 1.6 billion are second language (L2) speakers of English (Long, 2017). Because so much information is exchanged through written text, people around the world are required to read in more than one language. As a result, the rise of globalization has placed greater demands on people to become better at reading in a second language (L2). Given the large number of L2 speakers of English, this demand is especially important for these 1.6 billion English learners.

Researchers have suggested it is only through reading that L2 or foreign-language learners acquire the linguistic and topical knowledge required for improvement in reading

skills (Day & Bamford, 1998). Reading is a complex skill that requires intensive amounts of practice over extensive periods of time, but many L2 learners do not have enough time to adequately practice outside of the instructional setting in order to reach a proficient level (National Research Council, 2012). In foreign language settings, opportunities for extended reading experience are far more limited, making reading development extremely challenging for foreign language learners. As reading in the L2 takes on increasing importance, it is thus an important societal responsibility to provide all individuals with the opportunity to become a skilled and fluent L2 reader (Grabe, 2009).

During reading, lower-level and higher-level cognitive processes are simultaneously activated. Lower-level processes include lexical access or word recognition, syntactic processing, and semantic proposition formation (Ahmed & Han, in press). The reader processes letters, and then words, sentences, and finally the whole text to achieve meaning (Han, Anderson, & Freeman, 2009). Higher-level processes involve forming a meaning representation of the text (a text model of reading comprehension), drawing extensively on background knowledge to assign meaning to text, and making inferences – all of which require a considerable amount of cognitive resources. Both lower-level and higher-level processes are carried out simultaneously and very rapidly in working memory. However, working memory has limited capacity to perform multiple processes at the same time. Therefore, some processes (e.g., lower-level processes) need to be relatively automatic in order for both processes to work in parallel and efficiently. Through practice routines, such as repeated readings of texts, readers can automatize lower-level processes (e.g., word recognition skills), which allow them to re-allocate their attention to higher-level comprehension processing. The importance of efficient and

automatic word recognition skills as a prerequisite in successful and fluent reading should not be neglected in L2 reading (Day & Bamford, 1998); automatic word recognition skills are necessary for readers to be fluent in reading, and fluent L2 reading opens up the opportunity for people to excel in L2 development and the doors to the world.

Reading is not only an important literacy skill to develop, it also plays a crucial role in the development of second language proficiency. The field of Second Language Acquisition (SLA) has also embraced reading as an important focus of research. However, compared to the literacy paradigm described above, which has been largely influenced by first language (L1) research and which has treated reading as a literacy skill and thus focused on the role of comprehension, the SLA approach views reading as a significant source of target-language input that is necessary for the acquisition of an L2 (Han et al., 2009). SLA research argues that exposure to an abundant amount of meaning-bearing input (e.g., texts) is essential to L2 acquisition (VanPatten, 1996). Through reading practices, language instructors can provide crucial input sources to second and foreign language learners. Foreign language learners, in particular, have comparatively limited sources of input in the environment (Laufer, 2006), which makes L2 reading even more crucial for them (Han & Chen, 2010).

Fluent L2 reading not only requires one to have proficient reading skills, but it is also contingent upon a learner's linguistic competencies in a second language. Reading is a comprehending process in which we read to draw meaning from text, and it is essentially a linguistic process (Grabe & Stoller, 2011) during which we are also processing the text for linguistic form. However, this aspect of reading (i.e., reading as a linguistic process) is often downplayed (Grabe & Stoller, 2011). L2 reading research

generally agrees that fluent reading is dependent on a sufficient amount of L2 knowledge (i.e., vocabulary and grammar). Vocabulary knowledge, in particular, plays a major role in reading, because without it, nothing can be comprehended.

L2 reading research has highlighted a mutually dependent relationship between reading and vocabulary acquisition (Han & Chen, 2010). While reading provides a critical avenue of vocabulary development (Hudson, 2007; Zahar, Cobb, & Spada, 2001), the ability to read is also contingent upon a certain level of vocabulary knowledge (Han & Chen, 2010).

Researchers have suggested that vocabulary knowledge is "the most important component skill and resource for reading" (Grabe, 2009, p. 196) and a powerful predictor variable for text comprehension (Droop & Verhoeven, 2003; Qian, 2002). Jeon and Yamashita (2014) in their meta-analysis showed that L2 reading comprehension strongly correlated with L2 vocabulary knowledge, which underscores the importance of the role of lower-level processes and the building of a text-model of comprehension (Ahmed & Han, in press). Nonetheless, for adult language learners, the development of vocabulary and comprehension skills through direct instruction is particularly difficult (National Research Council, 2012). Even adult English language learners who are well educated "show weaknesses in vocabulary and comprehension" (National Research Council, 2012, p. 209). Because people need comprehension skills and vocabulary knowledge in order to have sufficient reading ability to respond to a variety of daily life, academic, and professional challenges, there is a need to investigate these concepts. Consequently, there is a strong demand for more efficient and effective ways of learning and teaching second language vocabulary and reading comprehension through L2 reading strategies. In

response, researchers and educators continue to seek out more effective approaches to promote L2 reading comprehension and vocabulary acquisition, and by extension, new avenues of L2 teaching and learning.

1.2 Theoretical Framework

Second-language reading is a complex process that can serve dual purposes – reading for comprehension and reading for language acquisition. SLA research has established that L2 readers use different sets of cognitive processes for comprehending informational content and for purposes of acquisition (Han & D'Angelo, 2009; Lee, 1998). SLA research has also evidenced that comprehension and acquisition are not synonymous (e.g., Gass, 2013; Gass & Selinker, 2008; Lightbown & Spada, 1990). This dissociation between comprehension and acquisition lends support to Sharwood Smith's (1986) conception of the dual relevance of information processing, where the learner interprets the input (i.e., information necessary for communicative exchange) for meaning and, at the same time, their acquisition device processes the linguistic input (i.e., aspects relevant to the learners' linguistic competence) for acquisition.

Such a dual purpose not only distinguishes L2 reading from L1 reading, but also it highlights the complexity of L2 reading (Han et al., 2009). First, unlike reading in the L1, second-language learners start reading in the target language before they achieve the level of grammatical and oral vocabulary ability that L1 learners achieve when they begin to read (Han et al., 2009; Shiotsu, 2009). Second, L2 reading research has suggested that the relationship between decoding and comprehension is more complex for L2 reading. L2 readers tend to struggle with word recognition skills and seldom achieve the same levels

of word-recognition fluency that are typical of L1 readers (Grabe, 2009). Also, L2 readers almost never achieve the same high level of fluency in reading comprehension skills as L1 readers (Grabe & Stoller, 2011).

Looking at the issue of language acquisition more closely, SLA theory has also established a fundamental difference between L1 and L2 acquisition (Bley-Vroman, 1989) concerning unique constraints and challenges that adult L2 learners confront. L1 learners generally acquire a language implicitly through experience with the language whereas L2 learners generally are less able to successfully learn through solely implicit means and therefore tend to need external assistance. Additionally, due to L1 interference (which may hinder the ability to notice linguistic features in the input) and maturational constraints (leading to lack of sensitivity to natural input), L2 learners are less susceptible to the benefits of mere exposure to L2 input. Even when provided with abundant L2 input (e.g., reading texts), learners may not be able to use that input to reconstruct their interlanguage system (Han & D'Angelo, 2009). Therefore, L2 learners need to rely more on additional support when learning a language (Han & D'Angelo, 2009).

Furthermore, according to research in cognitive psychology, humans have limited attentional capacity, which has implications for how learners process L2 texts. During input processing, meaning and linguistic form may compete for cognitive processing resources. Unlike L1 reading where processing of meaning and form happen simultaneously, L2 research has suggested that when focal attention to form occurs, it is often asynchronous with processing for meaning (Han & D'Angelo, 2009). Simultaneous processing of meaning and form can still occur, but only if the linguistic features carry significant information (e.g., lexical items).

Additionally, given that meaning takes priority for older learners and that learners' preoccupation is with constructing meaning from the text, a lot of what they read may fail to become useful input that is attended to and processed. Without attention to form in the input, learners will derive insufficient intake (i.e., the limited portion of incoming data attended to and processed by the learner for acquisition). Thus, learners should be provided with opportunities to attend to form while processing the text so that the input will generate intake and result in acquisition. Sharwood Smith's dual relevance theory implies that the input learners are exposed to needs to be sufficiently robust (i.e., consistent and frequent) as well as communicatively complex or diversified so that the input "may provide for both communicatively driven comprehension and acquisition" (Han & D'Angelo, 2009, p. 185).

L2 reading research and instruction has traditionally been based in either the literacy paradigm (which highlights the importance of comprehension) or the linguistic paradigm (which is primarily interested in acquisition), and the two paradigms have shown little crossover (Han et al., 2009). Researchers have come to acknowledge that an exclusion of or an overemphasis on either comprehension or acquisition processing results in a dissociation and imbalance between meaning and form and would most likely prevent true L2 development from taking place (Han & D'Angelo, 2009). As Han, Anderson, and Freeman (2009) explain, "a separation in instruction and research of the otherwise two interrelated dimensions is counterproductive to L2 learning" (p. 3). The concern should not be whether the focus of reading instruction should be on content or language. Rather, the question should be on developing comprehension and acquisition

hand-in-hand, as well as under what conditions learners can attend to both meaning and form at the same time (Han et al., 2009).

Acknowledging that both comprehension processes and acquisition processes are interdependent for L2 readers, there is a need to coalesce them within a single framework. In the history of L2 research and instruction there have been attempts to do just that. In the early 1980s, Krashen (1985) recommended that learners be exposed to a large amount of L2 comprehensible written texts as the means for achieving both comprehension and acquisition outcomes. This was called the comprehensible input model. However, rather than assuming that comprehensible input was all that is needed for acquisition to take place and that comprehensible input will lead to language acquisition, SLA research has empirically shown that acquisition does not necessarily follow as a byproduct of comprehension (Gass, 2013; Gass & Selinker, 2008). A focus-on-form approach (Long, 1991) tried to amend the comprehensible input model. Long's approach contends that incidental but overt attention to form within a meaning-based instruction may be beneficial for both comprehension and acquisition. Research within this paradigm, however, has found that meaning-based instruction combined with focus-on-form may still be insufficient for linguistic development (Han et al., 2009).

In a similar vein, building on Sharwood-Smith's dual relevance concept, Han and D'Angelo (2009) proposed a dual approach to teaching L2 reading, which aims at balancing both comprehension (i.e., development of a communicative ability) and linguistic competency (i.e., acquisition of form-meaning connections) within a meaningful context provided by written texts. In this way, the dual approach suggests a promising solution for L2 learner needs. The dual approach cultivates comprehension by

addressing reading efficiency needs and semantic processing needs through instructional

tasks that target reading skills such as skimming, scanning, and inferring meaning of

unknown words from the text, and reading strategies for comprehension such as top-

down and bottom-up support so that learners can read for a variety of purposes. At the

same time, the dual approach cultivates growth of linguistic competency by supporting

metalinguistic awareness and understanding of meaning-form connections. During

reading instruction, implicit and explicit strategies can be used to facilitate learners'

construction of relevant L2 linguistic knowledge. Furthermore, as Lee (2002) suggests,

learners would be able to make better meaning from the form and derive subsequent

intake when they are provided with opportunities to process the form in a meaningful

context through reading texts.

This study considers the needs of L2 learners through the framework of dual

relevance for L2 reading. In particular, it considers the condition of repeated reading,

described next, as a formative factor in promoting L2 reading comprehension and

incidental vocabulary acquisition.

1.3 Repeated Reading

Repeated reading (RR), an instructional approach devised by Samuels (1979) to

translate Automaticity Theory[1] into practice, involves "multiple, successive encounters

with the same visual material, the key being repetition – whether of the same words,

---

[1] An extension of the theory of automatic information processing in reading (LaBerge & Samuels, 1974), repeated reading was devised by Samuels (1979) to translate Automaticity Theory (Samuels, 1976) into practice. The essential role of automatic word recognition and fluency in reading is highlighted in Automaticity Theory, which posits that "a fluent reader decodes text automatically – that is, without attention – thus leaving attention free to be used for comprehension" (Samuels, 1979, p. 378).

sentences, or connected discourse" (Han & Chen, 2010, p. 243). In Samuels' original design, readers orally re-read a series of short passages from graded readers (i.e., books or passages with reduced range of vocabulary and simplified grammatical structures) until they reach a satisfactory level of fluency (i.e., accuracy of word recognition and reading speed), which is operationalized as being able to read at proficiency-appropriate word per minute (wpm) level. Upon achieving this level, the reader goes on to read the next passage, and the whole procedure is then repeated with a new passage. RR instruction can be implemented via various alternative procedures. For instance, the procedure can range from silent reading (Anderson, 1993, 2009), to reading aloud (Samuels, 1979), and to reading aloud while or after listening to an audio model (Chomsky, 1978). RR can be executed with the use of an audio model (assisted RR) or without (unassisted RR).

With its root in L1 reading research and instruction, RR was originally designed in response to the need to develop fluent word recognition skills and reading fluency in deficient L1 readers because these sub-skills are deemed to support comprehension. Research has provided abundant evidence suggesting the positive effects of repeated reading in improving L1 reading fluency, reading comprehension, and word recognition (National Reading Panel, 2000). However, it remains an empirical question as to its efficacy and role in L2 reading development and language acquisition.

Drawing on insights from L1 and L2 research, repeated reading has the potential to facilitate comprehension and acquisition in multiple ways. The following main potential benefits are worth noting. First, repeated reading provides the practice that is needed to develop automaticity in word recognition, which supports, but is not sufficient

for, reading fluency and subsequently reading comprehension overall. As a result of extensive RR practice, the reader is purported to free up most of their cognitive resources from lower-level processes (e.g., word recognition), which allows them to direct more attention to higher-order processes (i.e., comprehension of texts).

Second, repeated exposure to the same text or different portions of a longer text results in increased content familiarity, which in both L1 and L2 reading is deemed to facilitate comprehension of the reading passage (Han & D'Angelo, 2009). As the learners' comprehension increases, their attentional resources – which are preoccupied by semantic processing – could be freed up to attend to linguistic aspects and processing of form (e.g., vocabulary).

Third, through frequent exposure to the text, the recurrence of vocabulary items and linguistic forms will increase their saliency and increase the chance for them to be noticed, which according to the SLA literature is a necessary criterion for L2 acquisition (Han & D'Angelo, 2009; Schmidt, 2001). Given that meaning takes priority in older learners' input processing, one could assume that it is not a natural process to focus the learners' attention to form (Skehan, 1998). What is more, SLA research has evidenced that it is particularly hard for L2 adult learners to acquire linguistic items that are rare, semantically lightweight, less salient, and/or of little or no communicative value (Han & D'Angelo, 2009; Long & Robinson, 1998). By providing learners with some sort of pedagogic intervention – such as repeated exposure to the text through repeated reading – less obvious aspects of the input are possibly made salient and are noticed alongside that of meaning.

Finally, repeated reading provides the learner with repeated and frequent exposure to unknown words in meaningful contexts, which increases opportunities for the words to be learnt. In contrast to L1 vocabulary acquisition where the bulk of vocabulary is acquired incidentally through reading, L2 learners are less able to acquire words incidentally on their own and would need to rely on additional support. L2 incidental vocabulary acquisition through "normal" reading is a slow, incremental process that requires multiple exposures to the word in comprehensible context (Liu & Todd, 2016). Repeated reading, in some ways, can compensate for the limited exposure and resources (particularly in foreign language settings) by providing L2 learners with varieties of words and contexts and the frequent exposure cycles needed for acquisition of vocabulary.

1.4 Focus of the Study

Operating in the framework of dual relevance for L2 reading, the current study sought to bridge two disparate theoretical stances regarding L2 reading by investigating whether L2 repeated reading of texts promotes reading comprehension development and leads to vocabulary acquisition. The study also attempted to understand any possible interactions between comprehension and acquisition during repeated reading. Additionally, the study sought to address practical concerns by investigating the types of repeated reading intervention (assisted or unassisted) that lead to more optimal development of reading comprehension and vocabulary acquisition. To this end, the present study investigated an under-researched area in L2 research by examining the effects of repeated reading – unassisted and assisted – on both reading comprehension

and vocabulary acquisition for intermediate-level English-as-a-foreign-language (EFL) learners from northern Taiwan. While some studies have been conducted on the effects of repeated reading on either reading comprehension or vocabulary acquisition in the L2 setting, two major gaps remain in research.

For one, there is a scarcity in research on repeated reading in the L2 settings, not to mention flaws in the design of the studies (more detailed information will be discussed in the ensuing chapters). Abundant empirical evidence has suggested that repeated reading is a highly effective strategy for L1 reading. While findings for RR research in the L1 demonstrated converging evidence of its potential to enhance reading fluency and comprehension, RR research in the L2 (RR-L2) has received only scant attention. The very few existing RR-L2 studies have suggested some positive effects, but also mixed and inconclusive findings on the development of L2 reading comprehension or L2 vocabulary development through repeated exposure to the text. Furthermore, research generally suggested that dual-modality input is more beneficial than single-modality input for reading comprehension and vocabulary acquisition. The two RR-L2 studies that have teased apart and compared the dual-modality and single-modality conditions, however, come with limitations. Thus, more research is needed to see the results of the two types of intervention in RR-L2 settings.

Revealing whether RR is productive or counterproductive for L2 reading comprehension and vocabulary acquisition holds tremendous importance for L2 learners, in that reading is not only a literacy skill to be developed, but also a significant source of target-language input that is necessary for the acquisition of an L2. Additionally, confirming whether RR is beneficial for L2 reading comprehension and vocabulary

acquisition may advance the current understanding of RR as an intervention to promote L2 literacy learning and L2 acquisition of vocabulary.

Another notable gap in research is that the majority of extant RR studies have only investigated the effects of RR on comprehension and acquisition separately. Both should be empirically studied in tandem. Specifically, L2 studies have explored the effects of repeated reading for reading fluency and/or comprehension of EFL learners, vocabulary acquisition of a heritage speaker, or vocabulary acquisition of foreign-language learners. This is problematic and contradictory to SLA research findings, which have suggested that comprehension and acquisition need to be researched and instructed side-by-side in order to garner a complete picture of L2 learning and promote balanced, continuous, and effective L2 development (Han et al., 2009). Existing research has suggested that L2 RR leads to reading comprehension, and extant studies have also suggested that RR may lead to vocabulary acquisition. However, the findings are mainly from RR research examining the two separately, and little is known about comprehension and acquisition investigated in tandem. Therefore, the present study will contribute to this body of literature by examining the efficacy of RR on *both* reading comprehension and vocabulary acquisition for EFL learners. Only through examining comprehension and acquisition together can conclusions be drawn as to whether double treatment of reading comprehension and vocabulary acquisition is effective for L2 development, how the two can be brought together in L2 reading, as well as recommendations for implementing double treatment in a pedagogical setting.

1.5 Key Terms

The following terms are most relevant to the current study and are defined accordingly.

1. Repetition: In the context of RR, repetition is defined as multiple and successive encounters with the same reading text(s). Repetition occurs at multiple levels, e.g., word, sentential, and discourse levels.

2. Reading comprehension: Reading comprehension refers to the ability to understand written texts. Reading comprehension includes the process of simultaneously extracting, processing, and constructing meaning from written texts (Aebersold & Field, 1997; RAND Reading Study Group, 2002). It is operationalized as the amount of text information that is recalled.

3. Incidental vocabulary acquisition: Vocabulary acquisition is defined as knowledge of form-meaning mappings of words, and operationalized as the ability to recognize (i.e., receptive knowledge) and/or produce (i.e., productive knowledge) words in context and isolation (Han & Chen, 2010). In the case of incidental vocabulary acquisition, learners do not explicitly study (but may have inferred) the word meaning while reading; therefore, vocabulary acquisition is a by-product of reading, rather than a direct focus in its own right (Han & Chen, 2010).

4. Receptive vocabulary knowledge: Receptive vocabulary knowledge refers to learners' ability to recognize the L2 word form and provide the L1 meaning of the word.

5. Productive vocabulary knowledge: Productive vocabulary knowledge refers to learners' ability to produce the L2 word form from the L1 meaning provided.

6. Assisted repeated reading (assisted RR): With assisted repeated reading, the learner silently reads a passage along with the assistance of an audiotaped model of the text. With the use of an audio model, the learner is presented with both the phonic and graphic forms of the reading passage, and thus, receives both the visual input and the aural input of the text (i.e., dual modality input).

7. Unassisted repeated reading (unassisted RR): With unassisted repeated reading, the learner silently reads a passage without the use of an audio model. The learner is presented with the graphic form of the reading passage and receives only the visual input of the text (i.e., single modality input).

1.6 Outline of the Dissertation

The ensuing chapters of this dissertation are organized as follows. Chapter II reviews several strands of research that motivated the current study. These include prominent findings of previous RR research in L1 and L2 settings, focusing on the impact of RR on L2 reading comprehension, and L2 vocabulary acquisition through RR. The chapter concludes by summarizing categorical findings motivating the current study and identifying research gaps in the literature, and introducing the research questions of the current study. Chapter III presents the design and methodological procedures employed in this research. It begins with a description of the methodological issues of previous studies and an overview of the results of a pilot study that helped inform the methodology used in the current study. Then it discusses the design of the current study, with a description of the participants, the materials and target words, instruments and tasks, and procedures for data collection. The chapter ends with a discussion of the data coding and

data analysis procedures. Chapter IV reports on the present study's findings. Results from

the quantitative analyses are first presented, followed by a description of the

supplementary findings. Finally, Chapter V discusses the main findings in relation to the

research questions, to current research, and to our current understanding of the constructs.

The chapter concludes with a discussion of the limitations of the study and possible

directions for future research.

Chapter II

REVIEW OF THE LITERATURE

Several strands of research motivating the current study are reviewed in this chapter. First, this chapter will describe information processing in reading. This is followed by a review of previous repeated reading (RR) studies and prominent findings in the first-language (L1) and second-language (L2) literature, as well as a summary of categorical findings motivating the current study and identification of research gaps in the literature. Lastly, the research questions of the current study are presented.

2.1 Information Processing in Reading

An introduction of RR and what it entails was made in Chapter I. Before turning to a review of RR studies in the L1 and L2, this section will first discuss the most common underlying processes (e.g., lower-level and higher-level cognitive processes) that are activated during reading, and then discuss how repeated reading works to automatize the lower-level processes so that readers can direct their cognitive resources to higher-level comprehension processes.

Reading is a complex cognitive activity that involves a set of common reading comprehension processes. These processes are activated in working memory (Chang, 2012; cf., Daneman & Merikle, 1996 for a comprehensive review). Working memory refers to "the information that is activated, or given mental stimulation, for immediate storage *and* processing" (Grabe & Stoller, 2011, p. 12). According to well-established reading theories, working memory for reading encompasses the active and simultaneous

use of both lower-level and higher-level cognitive processes. Both processes play a critical role in reading. In lower-level (pre-lexical) processing, working memory supports three main processes: (1) lexical access or word recognition (i.e., identifying the meaning of a word as it is recognized in text) and processes that constitute word recognition, including letter feature extraction, orthographic segmentation, phonological coding and morphological processing for word recognition, (2) syntactic parsing, and (3) semantic proposition formation (or meaning proposition encoding) (Grabe & Stoller, 2011). The lower-level processes represent the more automatic linguistic processes, and they are generally considered as more skills-oriented (Grabe & Stoller, 2011). Skills instruction has traditionally been identified as a bottom-up approach to reading which emphasizes lower-level reading processes:

> Students start with the basics of letter and sound recognition, which, in turn, allows for morpheme recognition followed by word recognition, building up to the identification of grammatical structures, sentences, and longer texts. Letters, letter clusters, words, phrases, sentences, longer texts, and meaning is the order in achieving comprehension. (Han et al., 2009, p. 6)

Higher-level processing, or post-lexical access, speaks to the comprehension (or meaning construction) of the text – at the sentential, paragraph, and passage levels. Grabe and Stoller (2011) illustrate the higher-level comprehension process:

> The most fundamental higher-level comprehension process is the coordination of ideas from a text that represent the main points and supporting ideas to form a meaning representation of the text (a text model of reading comprehension) […] As clause-level meaning units are formed (drawing on information from syntactic parsing and semantic proposition formation), they are added to a growing network of ideas from the text. The new clauses may be hooked into the network in a number of ways: through the repetition of an idea, event, object or character; by reference to the same thing, but in different words; and through simple inferences that create a way to link a new meaning unit to the appropriate places in the network […] As the reader continues processing text information, and new meaning units are added, those ideas that are used repeatedly and that form usable linkages to other information begin to be viewed as the main ideas of the text.

More technically, they become, and remain, more active in the network. (p. 20) Readers comprehend and interpret the ideas represented by the text, draw extensively on background knowledge, and make inferences. In addition, higher-level processing also requires the ability to monitor comprehension, use cognitive and metacognitive strategies to understand text meaning, and critically evaluate the information being read, as well as establish purposes for reading, form attitudes about the text, task and author, assess and adjust goals as appropriate, and repair comprehension problems (Chang, 2012; Grabe & Stoller, 2011). Higher-level processing, in general, requires a considerable amount of cognitive resources (LaBerge & Samuels, 1974; Perfetti & Lesgold, 1976).

During reading, lower-level and higher-level processes are carried out simultaneously and very rapidly in working memory (Grabe & Stoller, 2011). While readers are recognizing words very rapidly and keeping them active in the working memory (Baddeley, 2007), they are also analyzing sentence structure to assemble clause-level meanings, constructing a main-idea model of text comprehension, and monitoring comprehension (Grabe & Stoller, 2011). "Combining these skills in an efficient manner makes general comprehension a time-consuming ability to master" (Grabe & Stoller, 2011, p. 11). Having to process these skills simultaneously is resource demanding, which makes mastering general comprehension difficult and time-consuming (Grabe & Stoller, 2011). Skilled readers are able to use both types of processing to engage in text comprehension and monitor whether or not the comprehension was successful (Gorsuch & Taguchi, 2010). However, a breakdown at any level of processing may jeopardize fluency and cause comprehension problems to occur (Grabe & Stoller, 2011; Logan, 1997).

Working memory is generally characterized as a limited-capacity system (Baddeley, 2006, 2007). It can hold a certain amount of information actively for only a very short period of time (Kintsch, Patel, & Ericsson, 1999) and has a limited ability to perform multiple processes simultaneously (Chang, 2012). Decoding and comprehension, however, compete for available cognitive resources (Grabe & Stoller, 2011). Thus, the dual process of word recognition and meaning construction is effortful and hard on memory (Samuels, 2006).

As Pressley (2006) explains, "when a reader slowly analyzes a word into component sounds and blends them, a great deal of capacity is consumed, with relatively little left over for comprehension of the word, let alone understanding the overall meaning of the sentence containing the word and the paragraph containing the sentence" (p. 68). If a reader expends too much attention on lower-level processing, less attention is available for higher-level comprehension processing (LaBerge & Samuels, 1974), and consequently, comprehension suffers. Additionally, a slow reader – often attending to every word and reading in isolated units (rather than meaningful sentences) – will likely have an overloaded short-term memory. Not being able to retain enough information in one's short-term memory for an extensive period of time is likely to render meaning construction processes of the connected text ineffective (Gorsuch & Taguchi, 2010).

Some processes need to be relatively automatic so that lower-level and higher-level processes can work in parallel and reading can work efficiently (Grabe & Stoller, 2011). Since lower-level processes are more amenable to automaticity training, it is logical to automatize the lower-level processing (i.e., word recognition) through practice,

e.g., repeated encounters of the text through repeated reading, so that learners are able to re-allocate their attention to higher-level processing in reading.

Drawing on insights from a general theory of automatic information processing in reading (LaBerge & Samuels, 1974), Samuels (1979) summarized three developmental stages of word recognition through which automaticity develops and comprehension is achieved. In the first stage, the non-accurate stage, learners have great difficulty in word recognition. They are non-automatic in decoding, and a significant amount of learner attention is required for word recognition. This in turn diverts attention away from the task of comprehension. Non-fluent readers achieve little comprehension and the reading process is laborious, most likely a result of the readers' poor word recognition skills (Gorsuch & Taguchi, 2008). For the second stage, the accuracy stage, word recognition is achieved with accuracy, but attention or cognitive effort is nevertheless required; comprehension at this stage is achieved, though it is very slow and halting. During the final stage, the automatic stage, word recognition is achieved. The reader is able to decode text automatically (i.e., without attention). As a result, the attention now becomes available for reading comprehension. For fluent readers at the automatic stage, they are equipped with fast and efficient processing abilities in order to achieve fluent reading comprehension.

In sum, fluent readers are characterized by their automatic word recognition and comprehension abilities, and non-fluent readers are characterized by slow decoding, which impedes flow of thoughts and comprehension. This highlights the importance of efficient and automatic word recognition skills as a prerequisite in successful reading (Day & Bamford, 1998; Grabe, 1991). That reading ability can be reliably determined

through the measurement of word recognition skills specifies the importance of automatizing one's word recognition skills so that the reader can achieve fluent reading comprehension.

An extension of the theory of automatic information processing in reading (LaBerge & Samuels, 1974), repeated reading was devised by Samuels (1979) to translate Automaticity Theory (Samuels, 1976) into practice. The essential role of automatic word recognition and fluency in reading is highlighted in Automaticity Theory, which posits that "a fluent reader decodes text automatically – that is, without attention – thus leaving attention free to be used for comprehension" (Samuels, 1979, p. 378).

At the heart of the repeated reading approach is *repetition* (Han & Chen, 2010; Perfetti & Roth, 1981). Through repeated exposure to the same text(s), repetition occurs at various levels (lexical, sentential, or contextual); additionally, readers' sight recognition of words and phrases will increase and automatize, which may further result in an increase in reading fluency, which is a necessary, but not sufficient, condition for improving reading comprehension (Blum et al., 1995; Dlugosz, 2000; Gorsuch & Taguchi, 2008; Taguchi, 1997; Taguchi & Gorsuch, 2002). In other words, when learners read a text repeatedly, they develop automaticity in lower-level processing (i.e., automatic, effortless word recognition skills) through practice. Automatic word recognition consumes little working memory capacity. As a result of extensive RR practice, the reader is purported to free up most of their cognitive resources, which allows them to direct more attention to higher-order processes; this results in improved comprehending and integrating of the word with the overall meaning of the sentence, paragraph, and text (Pressley, 2006; Taguchi, Gorsuch, & Sasamoto, 2006). Furthermore,

according to the theory of automatic information processing in reading (LaBerge & Samuels, 1974), repetition may lead to an increase in text familiarity and a decrease in the amount of information that needs to be processed (Hyönä & Niemi, 1990). Samuels (1979) concluded by suggesting an important function of RR in that it provides the practice needed to develop automaticity in word recognition skills, thereby contributing to fluency and comprehension.

2.2 Repeated Reading Research in L1

The repeated reading approach has been extensively studied in the L1 learning contexts. Despite various procedures implemented, RR research in the L1 has provided robust empirical evidence suggesting a variety of beneficial effects (for a comprehensive review of RR studies in L1 settings, see Kuhn & Stahl, 2003; National Reading Panel, 2000). The categorical findings include: (1) fluency and comprehension, (2) transfer of practice effects, (3) word recognition training in contexts, (4) larger and more meaningful, syntactically, and phonologically appropriate phrases (i.e., sentence segmentation and prosodic features), and (5) motivation and self-monitoring.

First and foremost, L1 research has reported beneficial effects of the RR approach for monolingual English readers' fluency and comprehension development. Empirical results suggest an increase in slow and halting readers' oral and silent reading fluency (i.e., reading rate and accuracy) through repeated exposure to written texts (e.g., Carver & Hoffman, 1981; Chomsky, 1976a, 1976b; Dahl, 1974; Dowhower, 1987, 1989; Faulkner & Levy, 1994; Herman, 1985; Young, Bowers, & MacKinnon, 1996). L1 research has also revealed a positive effect of RR on the decrease of word recognition

errors (Samuels, 1979) and on learners' vocabulary development (Elley, 1989; Koskinen & Blum, 1984). In addition, students show an increase in initial speed (i.e., the speed of first reading of each passage), and a decrease in the number of times a learner needs to read a passage in order to reach the satisfactory level of fluency (e.g., 85 wpm criterion rate) (Samuels, 1979). For example, while a reader initially required seven readings of the first passage in order to reach the criterion rate, he or she may progress to only needing three readings of the final passage to reach the satisfactory level of fluency; thus showing improvement in one's reading fluency through the repeated reading approach. Moyer (1982) in reviewing earlier L1 research concluded that RR assists general reading fluency "for some unskilled readers, for normal readers given difficult text, and in regular classroom instruction" (p. 620).

Research has shown that developmental benefits of re-reading a passage can go beyond reading fluency: the increase in reading rate and accuracy in turn leads to better comprehension of the reading texts (see for example Dowhower, 1987; Herman, 1985; Morgan & Lyon, 1979; National Reading Panel, 2000; Sindelar, Monda, & O'Shea, 1990; Young et al., 1996). However, whereas empirical evidence suggested that repeated reading increases both fluency and comprehension, Therrien (2004) in his meta-analysis revealed differential gains with respect to reading fluency and comprehension in RR. Such differential gains depend on whether the learner is reading the same or different passages, and whether the instructor cues the learners' attention to fluency and/or comprehension. From the meta-analysis, Therrien reported that repeated exposure to novel forms through re-reading the same text led to larger gains in fluency and comprehension than repeated exposure to novel forms through different texts.

Gains in reading fluency and comprehension from re-reading practice with one text have also been shown to transfer to new, unpracticed texts (i.e., transfer of practice effects) (Dowhower, 1989), albeit with some caveats. Samuels (1979), for instance, reported that learners' initial reading rates increased with each new passage and that fewer re-readings were necessary to reach a satisfactory level of fluency. This is an indication of transfer of training and a general improvement in reading fluency. Nevertheless, researchers have pointed out that a prerequisite for a high transfer of gains in reading rate to the new passage lies in a high degree of overlapping words between the old and new passages, which can be maximized by using a whole story or a series of related passages on the same topic (Rashotte & Torgesen, 1985).

The use of context is found to be crucial to facilitating successful RR in L1 settings. Studies have shown a preference for re-reading the entire passage and recognizing words in contexts (i.e., connected discourse), as opposed to in isolation. There is empirical evidence suggesting that word recognition training in isolation (e.g., re-reading of word lists) does not lead to better comprehension of the text containing the target words (Dahl, 1974; Fleisher, Jenkins, & Pany, 1979). On the other hand, other researchers argued that in repeated reading, the practicing of one's word recognition skills in context may actually facilitate one's reading comprehension (Dahl, 1974). Moyer (1982) expresses this idea in her article:

> … repeated reading of entire passages makes it possible for readers to maximise the redundancy found in written language. That is to say, using the context to recognise words in repeated readings, readers may also become familiar with the semantic and syntactic redundancy in the text. By capitalising on this semantic and syntactic information, readers become able to comprehend some words in chunks. Reading in large, meaningful units is a characteristic of fluent readers and serves to facilitate better comprehension (e.g., Cromer, 1970; Smith, 1978). (Taguchi, 1997, p. 101)

Furthermore, RR allows learners to read in larger and more meaningful, syntactically, and phonologically appropriate phrases (Dowhower, 1987). "Meaningful" refers to one's ability to segment sentences into larger meaningful units. Through re-reading of a passage, often combined with guided phonological practice, readers may gradually learn the prosodic features (e.g., intonation, stress, duration, and pause) of a language and generate appropriate prosodic markings, which will allow learners to accurately segment sentences into meaningful phrases (Han & Chen; 2010; Schreiber, 1980). By doing so, readers can aid fluency and comprehension.

Advocates of L1 repeated reading also suggested that it motivates students to read more. An essential element for generating learner motivation is the visual depiction of their gains in reading rates and decreases in reading errors represented in a graph. In the RR approach, a learner times and records his or her reading speed and miscues for each consecutive reading exercise throughout the whole treatment (Samuels, 1979). Once the learners see the graph, the learners' progress throughout the course becomes evident and thus sparks their motivation. Empirical and pedagogical evidence has shown that the RR method seems to motivate students, rather than making them bored by reading the same passage over and over again. Chomsky (1978, as cited in Han & Chen, 2010) reported that RR instilled in the reader "a heightened sense of confidence, motivation, and willingness to undertake reading new material independently" (p. 2). It appears that repeated reading enhances self-monitoring, an essential behavior for independent readers (Blum et al., 1995), and that it helps learners gain confidence in their reading (Koskinen & Blum, 1984).

2.3 Repeated Reading Research in L2

Whereas repeated reading has received considerable attention and revealed positive converging evidence in L1 research, RR is a rather new instructional approach in second- and foreign-language reading contexts (Taguchi, Takayasu-Maass, & Gorsuch, 2004) and has received only scant attention in L2 research. One reason may be due to a prevailing assumption among L2 educators that there is a positive correlation between one's reading fluency and overall language proficiency, such that the former grows as the latter improves (Gorsuch & Taguchi, 2008). This assumption, nevertheless, has received little empirical support (Koda, 2005). Rather, empirical evidence has suggested the opposite: increase in language proficiency does not necessarily ensure good reading fluency (Favreau & Segalowitz, 1983; Segalowitz et al., 1991). Another reason may be due to practitioners' lack of attention to fluency, which may have resulted from the lack of understanding of the role that automatic word recognition plays in reading (Grabe & Stoller, 2011). To carry out reading comprehension for an extended period of time requires the readers to have good word recognition skills, which are difficult to develop without exposure to print through extensive reading practice (Grabe & Stoller, 2011). This lack of awareness (i.e., the essential role word recognition plays in reading fluency and comprehension) may also be a consequence of a prevailing attitude towards reading as a vehicle for the study of grammar and lexis, test preparations, and oral proficiency (Gorsuch & Taguchi, 2010). Gorsuch and Taguchi (2010) noted that this is particularly the case in foreign language educational settings. One other reason that fluency development has been ignored in the past may be due to the difficulties in providing sufficient amounts of time, resources, and practice opportunities – which are much more

limited in foreign language settings – that are required for developing word recognition skills (Gorsuch & Taguchi, 2010) and a large sight vocabulary (Grabe & Stoller, 2011). In general, foreign-language instruction has not been constructed in a way that may support reading fluency and comprehension development.

As fluency development became a significant pedagogical issue in L2 reading in the past decade (Gorsuch & Taguchi, 2008; Grabe, 2004; Grabe & Stoller, 2011; Nation, 2001), so too has RR received increasing attention in research in the L2 reading contexts. Similar to RR research in the L1 settings, accumulated empirical evidence has contributed to the notion that RR can be effective for L2 readers. Most prominently, it may be a promising method for improving L2 fluency, which leads to better comprehension (Grabe, 1991).

2.3.1 Reading Fluency

To date, only several empirical studies have directly measured the effectiveness of RR on the reading fluency and/or comprehension of second- or foreign-language adult learners (e.g., Chang, 2012; Chang & Millet, 2013; Gorsuch & Taguchi, 2008, 2010; Gorsuch, Taguchi, & Umehara, 2015; Taguchi, 1997; Taguchi & Gorsuch, 2002; Taguchi, Gorsuch, Takayasu-Maass, & Snipp, 2012; Taguchi et al., 2004). The extant studies have yielded positive, albeit inconclusive, findings.

First of all, results from quantitative and qualitative data have suggested that RR is an effective approach for developing foreign-language word recognition skills (e.g., Taguchi, 1997; Taguchi & Gorsuch, 2002) and for increasing reading rate and reading fluency both within each session and across the treatment sessions (e.g., Chang, 2012; Chang & Millett, 2013; Gorsuch & Taguchi, 2008; Taguchi, 1997; Taguch & Gorsuch,

2002; Taguchi et al., 2012; Taguchi et al., 2004). Whereas there was a significant increase in reading speed from the pre-test to the post-test in Chang (2012), Chang and Millett (2013), Taguchi and Gorsuch (2002), and Taguchi et al. (2012), other studies did not show existence of such transfer of practice effect in terms of reading rate to a new, unpracticed passage (i.e., a passage different from that used for the intervention).

Chang (2012) in her study compared the effects of two distinct interventions – timed reading (TR; n=18) and repeated oral reading (RR; n=17) – on Taiwanese EFL learners' reading rate, reading comprehension, and learner perception. To test the effects of the treatment on the learners' reading rate and comprehension, a pre-test, post-test, and delayed post-test (six weeks later) were administered. Results suggested that both the TR and RR groups' reading rates improved significantly to varying degrees after the intervention, but the TR group improved significantly more (i.e., twice as much) than the oral RR group. For both groups, the increase was retained six weeks after the intervention with only marginal attrition. According to Chang (2012), such differences between the two groups' gains in reading rate may have been due to factors such as a less amount of reading required for the RR group (i.e., the RR group read only one out of three passages read by the TR group); according to extensive reading studies, increased exposure to print increases reading rates. In addition, in contrast to L1 studies, the participants did not consider oral re-reading a method for improving silent reading fluency; as reflected in their report, they considered it an approach for improving oral production and pronunciation (Chang, 2012). The differences between the RR and TR groups suggest that the two approaches seem to focus on training of different aspects (e.g., the RR method more on accurate and automatic word decoding and chunking of texts, and the

TR more on training reading skills and reading for general comprehension) and thus, they are not comparable in many ways (Chang, 2012). Another influential factor may be related to the design of the treatment: the repeated oral reading group not only performed three oral re-readings combined with one assisted reading and one unassisted silent reading, but also received audio assistance, pronunciation correction, miscue feedback, and peer interaction. Without such additional support, the RR group's improvement may have been smaller (Chang, 2012).

In a more recent study, Chang and Millett (2013) investigated the effects of timed repeated reading on reading rates, comprehension level, and potential transfer effects. They compared the repeated reading group (RR, n=13) to the non-repeated reading group (NRR, n=13). For the treatment sessions, the RR group silently read two expository passages (approximately 300 words per text) five times each, and after the first reading and the fifth (final) reading, they answered the same comprehension questions measuring global understanding. The NRR group read the same passages but without repetition and answered the comprehension questions. For the pre-test and the post-test, all participants read two practiced passages and an unpracticed passage (i.e., a 2058-word narrative from a graded reader) once, and then completed the comprehension questions. After the intervention, the RR group improved 47 wpm (from 103 to 150 wpm) or 46% on the practiced passages and 45 wpm (100 to 145 wpm) or 45% on the unpracticed passages; on the other hand, the NRR group only improved 13 wpm (107 to 120 wpm) or 12% on the practiced passages and 7 wpm (102 to 109 wpm) or 7% on the unpracticed passages. Results suggested better improvement of reading rate for students who read the same passages multiple times than for those who read only once, with the RR group reading 30

and 36 wpm faster than the NRR group on the practiced and unpracticed texts, respectively. Results also suggested that the effect gained from the practice could be transferred to an unpracticed passage. Whereas the difference between the RR and NRR groups' initial reading rate was not significant for the pre-test, the difference was found to be statistically significant for the post-test for both the practiced and unpracticed passages.

2.3.2 Reading Comprehension

Similar to studies in L1 settings showing converging evidence of the positive effect of RR on reading comprehension, results from existing RR-L2 studies suggested converging evidence that RR led to reading comprehension of the same text(s), but mixed findings in terms of a transfer of practice effect to new, unpracticed passages.

In an earlier study, Bernhardt (1983b) compared multiple silent and oral readings and the relationship among reading mode, grammatical ability, comprehension, re-reading, and perception of text difficulty. Fourteen intermediate-level L2 German learners (in their fourth-semester German classes in a US college) read two expository texts describing basic psychology experiments. Half of them were categorized as good and half as poor on their grammatical ability based on the AATG Standardized Achievement Test – Level IV. The participants read both Passage A (139 words) and Passage B (134 words): they either read Passage A silently and Passage B orally, or vice versa. Some participants (the author, again, did not specify the number) read Passage A first, while others started with Passage B. The participants read the same text three times and were notified that they will recall the passage in writing in their L1 English after every reading. The texts were divided into constituent structures (Passage A: 55

constituent phrases; Passage B: 52 constituent phrases), which consisted primarily of nouns, verbs, and prepositional phrases, and were coded for appearance of a meaningful phrase. Significant main effects were found for grammatical ability (good or poor), mode (oral or silent reading), and repetition (first, second, or third reading). Students with good grammatical ability ($\overline{X}$=23.69) had significantly higher comprehension scores than students with poor grammatical ability ($\overline{X}$=15.36). Comprehension of passages read silently ($\overline{X}$=22.47) was significantly higher than that of passages read orally ($\overline{X}$=16.58), for each level of repetition and regardless of grammatical ability. Almost all participants perceived the text that was read orally to be more difficult, regardless of the passage read. Comprehension improved significantly from the first ($\overline{X}$=14.09) to the second ($\overline{X}$=20.17) and to the third ($\overline{X}$=24.31) reading for all participants, regardless of their grammatical ability. This implies the benefits of re-reading and that re-reading, as Bernhardt (1983b) suggests, should play an important role in foreign language instruction. However, while the participants continued to show improvement with each successive reading, the study was not in the position to make any claims for performance beyond the third reading (Bernhardt, 1983b). Further investigation to identify the optimum number of reading necessary for the participants to hit the point of diminishing returns and achieve utmost outcome under different conditions should be considered.

Motivated by Taguchi (1997) which examined whether or not RR is an effective method for beginning-to-intermediate-level EFL learners in Japan in increasing their silent and oral reading rates, Taguchi and Gorsuch (2002) and Taguchi et al. (2004) investigated the effect of the RR treatment on silent reading rate and comprehension, and the transfer effects of reading gains to new texts. The Japanese university participants

were divided into the RR and the extensive reading (ER) groups. Both the RR and ER

groups showed significant increase in their comprehension scores on the pre-test and the

post-test as the number of re-reading increased within the session; however, no transfer

effects from the pre-test to the post-test passages were found in terms of reading

comprehension, and no significant differences were found between the RR and ER

groups (Taguchi & Gorsuch, 2002; Taguchi et al., 2004). Taguchi and Gorsuch (2002)

have found that the passages chosen for the pre-test and the post-test sessions were rather

difficult for the participants to begin with (i.e., the participants scored 35% correct on the

comprehension pre-test and 50% correct on the post-test), and were thus, not sensitive

enough to reveal changes from the treatment. In their study, the RR group's initial

reading rate on the post-test (153 wpm) was considerably lower than the minimum rate

(i.e., 200 wpm) for readers to read with full comprehension, implying that the readers

were not able to free themselves from word recognition tasks, to process the text

efficiently, and to use appropriate higher order comprehension skills that would facilitate

better performance in comprehension (Taguchi & Gorsuch, 2002). Another explanation

for the insignificant difference between the RR and ER groups, as Taguchi and Gorsuch

(2002) argue, could be that the treatment passages were not designed for the purpose of

fluency training, and therefore, the vocabulary and syntactic complexity were not

controlled for.

A different pattern was shown in Gorsuch and Taguchi (2008). They found that

both the RR and control groups improved from the initial reading to the fifth reading on

both test conditions (i.e., the short answer and the recall tasks). The RR group showed

significant increase on their post-test scores; the control group, on the other hand, did not

show an increase from the pre-test to the post-test. Even though both groups performed similarly in terms of their post-test reading rate, the RR experimental group comprehended significantly more than the control group on both the first and fifth readings for the post-test tasks. As Gorsuch and Taguchi (2008) speculated, the RR group took better advantage of the repetition during the post-test: because the RR group knew they would be answering comprehension questions after reading, they read more slowly and more carefully in order to do well on the comprehension post-tests, suggesting use of metacognitive strategies. Nevertheless, the comprehension levels remained low, not reaching the 70% satisfactory level suggested by Nation (2005).

Aside from Gorsuch and Taguchi's 2008 study, also part of the same larger study was Gorsuch and Taguchi's (2010) longitudinal, qualitative study on 30 Vietnamese intermediate-level EFL college students. Analyses of open-ended student reports revealed that some of the participants were aware of the trade-off between reading speed and comprehension. They also commented that the fluency and comprehension building effects that generated from the RR method transferred, or "spilled over," into their reading outside of class. A number of participants observed their increases in comprehension and reading speed, and some of their comments particularly attributed the effect to the RR treatment. The participants' perceived improvement in comprehension and speed was also corroborated by their score increases demonstrated by the 2008 quantitative data.

Similarly, Taguchi et al.'s (2012) longitudinal diary case study of an advanced-level Japanese EFL reader, Naomi, revealed an increase in both reading speed and comprehension scores from the pre-test to the post-test, suggesting a transfer effect of re-

reading practice to a new, unpracticed passage. As Naomi's fluency increased with re-readings of a text, she was able to free up her cognitive resources (from lower-level processing) to achieve nuanced comprehension of the text. The qualitative data constituted post-session diary entries that tapped into metacognitive processes, from which two recurring themes emerged: beneficial and non-beneficial effects of RR on comprehension, as well as the use of an audio model (which will be discussed later in this chapter). Naomi's comment suggested that RR improved overall comprehension of the texts, as well as comprehension at the sentential and discourse levels. In addition, RR provided scaffolding in three various ways, including: (1) allowing her to identify incomprehensible parts of the text; (2) assisting her in determining the importance of unfamiliar words or phrases and in guessing meanings of unknown words, phrases, or grammatical ambiguities through contextual support; and (3) helping her with retaining and productive use of words and with retrieving word meaning. The aforementioned benefits, shown from quantitative and qualitative data, would not have emerged if the text were read for the first time and read only one time, where the reader may still be struggling with word recognition and parsing processes and where the reader's cognitive resources are likely to be expended on lower-level processes which remain resource-demanding for him/her and, as a consequence, inhibiting higher-order comprehension processes (Taguchi et al., 2012). On the other hand, Naomi commented that too many re-readings of the same passage led to boredom and demotivation and caused her to lose concentration. Furthermore, RR alone did not make incomprehensible points (e.g., unknown words) comprehensible through contextual support, which led to frustration. As the researchers speculated, if there were many unfamiliar words and phrases, the reader

may not be able to guess the meaning from limited information available in the context, and thus, RR alone may not lead to better comprehension.

An interesting discussion was raised by Taguchi et al. (2012) in that they speculated that Naomi was actually engaged in "reading to learn," which involved not only understanding, but also remembering the ideas in the text (Fraser, 2007). This requires a deeper processing and is different from "rauding," where the reader integrates and quickly processes ideas in sentences (Carver, 1990). When L2 learners read a text to learn, their reading rate slows down to 200 wpm or less (Fraser, 2007). Thus, that Naomi was actually engaged in reading to learn might be a reason why her mean rate for the initial readings of the treatment texts was comparatively slow, with only 131.44 wpm (*SD* = 15.10); this can be reflected by her constantly questioning semantic and syntactic ambiguities in the text so as to achieve high-level comprehension.

A different pattern was found in Chang's (2012) study comparing the effects of timed reading (TR; n=18) and repeated oral reading (RR; n=17) on Taiwanese EFL learners' reading rate, reading comprehension, and learner perception. For the treatment sessions, the RR group read a total of 26 segments, with 300-400 words per passage; they read one of the three passages that were read by the TR group. The procedures for the RR group were as follows; they read the passage once with an audio model, silently read the passage once, answered comprehension questions and checked the answers, orally read the passage twice individually, orally read the passage once in pairs, and lastly, one of the participants volunteered to read in front of the whole class and the instructor provided feedback. In this sense, the RR group participants were engaged in one assisted reading, one silent reading, three oral readings, in addition to being provided with assistance on

pronunciation and meaning of the text when needed, provided with miscue feedback, and engaged in peer interaction. On the other hand, the TR group read a total of 52 segments, with 300-400 words per passage; they read three passages per session, and following timed reading of each passage, the participants completed comprehension questions. For the pre-test, all participants read three passages, each consisting of about 400 words; the same passages were used for the post-test. For the delayed post-test, three 400-word passages – different from the ones used for the pre-test and the post-test – were provided.

As reported by Chang (2012), the TR group improved 14% on the comprehension post-test and retained 10% six weeks later, whereas the RR group's comprehension scores did not demonstrate significant variability across the three tests. Also, the TR and RR groups did not differ significantly after receiving different interventions. Chang suspected this may be because the students were still unable to process the lower-level components (e.g., word recognition) automatically and could not direct their attention to comprehending the text. For both groups, their comprehension scores did not show a decline as their reading rate increased. As Chang (2012) pointed out, this suggested that focusing on improving one's reading rate does not have a detrimental effect on reading comprehension. Qualitative data were also analyzed, as the participants' perception of the task goal could have a strong effect on the participants' focus during task performance (Chang, 2012). Results revealed positive opinions of the intervention: the RR group perceived their improvement in fluency (59%) and pronunciation (53%) and outlined an increase in their concentration (41%) and confidence (18%). Also, the RR group participants expressed that they felt tired after the fourth or fifth re-reading and were frustrated when encountering hard-to-pronounce words (29%).

Conversely, Chang and Millett (2013) found that while the RR and non-RR (NRR) groups scored similarly on the comprehension pre-test (51% and 49% comprehension levels, respectively, for the practiced passages; 49% and 44%, respectively, for the unpracticed passage), the RR group improved much more than the NRR group on the comprehension post-test (the RR group: 70% comprehension level, or 19% improvement for practiced passages, and 66% comprehension level, or 17% improvement for the unpracticed passage; the NRR group: 54% comprehension level, or 5% improvement for the practiced texts, and 34% comprehension level, or 7% improvement for the unpracticed text). Also, the effect was transferred to the unpracticed passage. While the RR group scored higher than the NRR group on both the practiced and unpracticed passages, a significant difference was only found in the unpracticed passage. Interestingly, an informal post-study interview with seven RR-group students discussing their perceptions of the treatment suggested that recording their reading time helped them concentrate better. Due to time pressure, they could not stop whenever to look up unfamiliar words and had to guess the meanings from the text. Two of the students commented that it was boring reading the same passage five times and suggested that three re-reading would be sufficient.

More recently, Gorsuch, Taguchi, and Umehara (2015) investigated the use of RR for fluency and comprehension building for beginning-level Japanese-as-a-foreign-language learners in a US undergraduate college. The study consisted of 23 treatment sessions over one semester; the learners participated in the study twice a week, 20 minutes per session. They read contemporary stories and folktales taken from a mainstream Japanese elementary school language arts textbook. For each treatment

session, the participants read one segment of a story, which was approximately 79

characters, or 30 words (range: 36 to 107 characters, or 16 to 43 words); one illustration

of the content was included on almost every page. Each passage was read a total of five

times: the participants silently read a segment of a story once and were invited to ask

questions about the text (e.g., vocabulary); then they read the text along with an audio

model two times, and silently read the text another two times. The participants timed their

reading rate for the first, fourth, and fifth readings. Lastly, they wrote a short open-ended

report (in their L1 English) reflecting on what they were thinking as they were reading,

how much they comprehended, and whether there were changes in their thinking, reading,

or comprehension.

For the testing sessions, the same reading procedure was implemented but without

the invitation to ask questions portion and the open-ended writing task. For the pre-test, a

148-character, 58-word passage was used; the same passage was used for the post-test.

The testing passage was slightly longer than the treatment passages. The participants

were measured on their reading rate for the first, fourth, and fifth reading. After re-

reading the testing passage five times, a recall protocol procedure (i.e., recall in their L1)

was used to gauge their reading comprehension. The pre-test and the post-test took place

three months apart to "minimize learners' memory-based knowledge of the text content"

(Gorsuch et al., 2015, p. 25).

Results suggested that the learners showed an increase of the hiragana character

(which is mainly used for function words to show case, adjective inflections, and adverb

inflections) and word decoding skills, as well as improvement on comprehension

(Gorsuch et al., 2015). With regards to the participants' reading rate, there was an

increase in reading rate within RR sessions (i.e., first, fourth, and fifth re-readings); additionally, the participants' initial reading rates increased over time. The researchers speculated that the learners used more automatized character and word decoding, and word recognition skills on new, unpracticed passages. In terms of their comprehension scores, the participants improved from an average score of 6.09% on the pre-test (range: 0% to 21%) to an average of 33.11% on the post-test (range: 14% to 50%). The participants also reported to have used a variety of comprehension strategies and to have more confidence in their abilities to read in Japanese texts that are easier and authentic. As Gorsuch et al. (2015) speculated, the participants' improvement in hiragana character and word recognition skills seemed to have contributed to their better word meaning recognition and therefore led to better comprehension.

The studies reviewed in this section seem to indicate that L2 repeated reading led to improvement in reading comprehension of the same text. However, findings were less conclusive as to whether L2 repeated reading can further lead to a transfer of practice effect to new, unpracticed passage in terms of reading comprehension. Thus, more research is needed to lead to a decisive finding.

2.3.3 Vocabulary Acquisition

Research has suggested that reading can not only promote reading comprehension, but also promote L2 vocabulary growth. While the aforementioned studies looked at whether RR (with its root in the L1 literacy paradigm and its original goal to facilitate reading fluency) is effective in facilitating reading fluency and comprehension, none of the studies addressed the issue of how RR can be helpful in L2 acquisition. As explained in Chapter I, SLA research has established that comprehension does not equal acquisition

and that acquisition does not necessarily follow as a by-product of comprehension (Gass, 2013; Gass & Selinker, 2008). This underscores the need for systematic research in the acquisition domain. In the recent years a few studies have investigated L2 vocabulary acquisition through repeated reading (e.g. Han & Chen, 2010; Liu & Todd, 2014, 2016; Webb & Chang, 2012; Zahar et al., 2001).

In L1 vocabulary acquisition, the bulk of vocabulary is acquired incidentally through reading. However, L2 reading is different from L1 reading, particularly in terms of vocabulary coverage required, vocabulary acquisition rate, and lexical access (Cobb, 2009). A few studies have examined the percentage of the vocabulary that needs to be known for L2 learners to understand the discourse and its effect on the level of comprehension achieved. For instance, Hu and Nation (2000) in their study investigated unknown word density and reading comprehension – as measured by a multiple-choice test and a cued written recall test. They found that in order for English language learners to achieve reading comprehension of fiction texts, 98% known word density is required to achieve adequate comprehension for most learners. When known word density lowered to 95%, a small minority of participants was able to gain adequate comprehension (Hu & Nation, 2000; Nation, 2006). With 90% known word density, an even smaller minority achieved adequate comprehension. In another study, Laufer and Ravenhorst-Kalovski (2010) investigated lexical text coverage, vocabulary size, and reading comprehension of Israeli EFL learners. Similar to Hu and Nation's (2000) study, Laufer and Ravenhorst-Kalovski (2010) found 98% as the sufficient and 95% as the minimal threshold to achieve "adequate" reading comprehension. 98% known word density enables adequate reading comprehension and independence, and successful

comprehension by most learners. With 95% known word density, learners can reach "acceptable" reading comprehension, which is probably viable with some support (e.g., instruction or resources). Likewise, Schmitt, Jiang, and Grabe (2011) found that with 98% text coverage, the reader would have 60% reading comprehension of the text. On the other hand, van Zeeland and Schmitt's (2013) study examined lexical coverage required for listening comprehension. The participants were required to listen to each story two times, and a multiple-choice test was given for each anecdote to measure their listening comprehension. At the 95%-98% known word density level, the participants demonstrated acceptable comprehension. Van Zeeland and Schmitt also found that listening comprehension was still quite good and good enough at the 95% and 90% levels, but learner performance was more consistent at the 95% level and showed higher variability at the 90% level. According to the results from extant research, it seems that "about 98% is the lexical coverage which is most appropriate for most purposes involving written text" (Schmitt, Cobb, Horst, & Schmitt, 2017, p. 215). Under most circumstances, 95% known word density level may still be sufficient for acceptable reading comprehension (Schmitt et al., 2017). Research in general has suggested that if the learner cannot read at least 95% of the words, comprehension will be impeded (Klingner, 2004).

With respect to vocabulary acquisition rate, L2 incidental vocabulary acquisition through reading is a slow, incremental process that requires multiple exposures to the word in a comprehensible context (Liu & Todd, 2016). The literature has supported that to increase word retention rate, the learner should be provided opportunities to repeatedly encounter the words; likewise, L2 research has suggested an even greater positive effect

if the words are recycled in different ways or contexts and in successive levels of depth (Thornbury, 2002). Research has proposed different numbers of word repetition necessary for successful vocabulary acquisition, but the general consensus appears to be that at least six or seven encounters are required for most vocabulary items to be learned (Rott, 1999), although in some cases five encounters may allow learning to occur. Nonetheless, L2 learners in general do not read extensively enough to reach the amount of reading needed to be able to incidentally acquire even the most frequent 3000 words (Cobb, 2007; Liu & Todd, 2016). Even when incidental acquisition does occur, the vocabulary gains seem to be very limited through comprehensible contextual input alone. Horst, Cobb, and Meara's (1998) study, for instance, showed an average of five words acquired incidentally after exposure to a 21,000-word comprehensible text. In repeated reading, on the other hand, a slightly different and more promising picture emerges (Liu & Todd, 2016).

For example, Han and Chen's (2010) case study attempted to gauge the type and depth of vocabulary acquisition (i.e., knowledge of form-meaning mappings of words) by a heritage speaker of Mandarin Chinese through RR-based instructional strategies. During the twenty sessions over a three-week period, the participant was engaged in assisted RR until fluency was achieved so that her attentional resources could later be allocated to higher-order processing, namely comprehension. Then, repeated reading continued to take place, interwoven with explicit form-focused instruction, including vocabulary enhancement activities, corrective feedback on word recognition errors and at the sentential level, writing practices, and comprehension check. Special assistance for both meaning and vocabulary acquisition (i.e., form) was provided to the participant

when needed. Results showed that RR combined with various pedagogical components promoted the participant's motivation to learn and her comprehension of authentic reading materials. Moreover, frequent and consistent repeated encounters of the same texts led to positive but differential vocabulary gains of intentional and incidental words in context and in isolation (i.e., contextual words > isolated words, word recognition > word production, and intentional words > incidental words). Words that received conscious attention and that were provided with corrective feedback were better retained. The positive gains suggested that RR could become an interesting source of vocabulary acquisition. In addition, RR allowed the participant to process texts that were substantially beyond her decoding capacity and level of processing or comprehension (Han & Chen, 2010). This corroborates with previous RR research findings in L1, documenting the potential of RR for difficult texts (Hyönä & Niemi, 1990; Moyer, 1982). The participant's comprehension was checked by asking the participant to explain her interpretation of the text. The participant demonstrated the ability to critically comment on logical inconsistencies of the texts, which is an indicator that the participant's comprehension was outstanding and that she was able to process texts that were far beyond her capacity to decode. Nonetheless, since the main focus of Han and Chen's study was on vocabulary acquisition, comprehension was not systematically measured; therefore, how much understanding was actually achieved and promoted through RR remains unknown. Furthermore, it must be noted that since the RR treatment is combined with other strategies, the positive gains reported in Han and Chen's study may not be attributed to RR alone and that other instructional strategies used would most likely have contributed to the positive results.

A similar positive finding was found in another study investigating RR and incidental vocabulary acquisition. Webb and Chang (2012) studied the effects of assisted RR (i.e., silent reading and listening) and unassisted RR (i.e., sight reading) on 82 beginner adolescent EFL readers' incidental vocabulary learning. Over two seven-week periods of study, the participants met once a week and read a total of 28 short stories. For each class, they read two short stories (approximately 300 words per passage), with each text being read a minimum of two times. Of the 100 target words, 85 of them appeared once and 15 appeared at least twice in the text. To measure the participants' form-meaning mappings of the words, the researchers administered a modified Vocabulary Knowledge Scale (VKS) for the pre-test and the post-test. Their study showed that beginner readers may acquire word meanings incidentally through assisted RR (12.03% and 14.14% vocabulary gains in the first and second periods of study, respectively) and unassisted RR (4.41% and 5.90% vocabulary gains, respectively), and that assisted RR led to greater incidental vocabulary gains than unassisted RR. However, such vocabulary gains cannot be solely attributed to RR, as the participants were allowed to resort to the dictionary, raise questions, or discuss the content with their peers during the treatment sessions, which may have promoted vocabulary learning. Furthermore, the study did not include a control group, which may have diluted its assertion that the gains were solely due to the treatment.

While the Han and Chen (2010) and Webb and Chang (2012) studies provided encouraging results for RR in facilitating vocabulary acquisition, they do not speak to the development of vocabulary through solely RR. On the other hand, a few studies have

investigated incidental vocabulary acquisition through RR – without the inclusion of a more explicit form-focused instruction or other external assistance.

In Zahar, Cobb, and Spada (2001), 144 English-as-a-second-language (ESL) learners were placed into five proficiency groups (ranging from beginner to bilingual), and the participants re-read a graded reader along with an audio model. Their incidental vocabulary acquisition (i.e., mapping contextually-appropriate meaning to the form of the words) was measured using a multiple-choice word definition test. The results suggested that incidental vocabulary acquisition occurred after an average of seven encounters of the target words in the text. Zahar, Cobb, and Spada also indicated that proficiency played a role in the successfulness of incidental vocabulary acquisition – where higher-level proficiency learners were able to acquire the words incidentally with fewer encounters of the words in context.

Liu and Todd (2014) investigated the efficacy of dual modality input on L2 learners' reading comprehension and incidental vocabulary gains through RR. The efficacy of the treatment was also examined in relation to learners' preferred learning style. Eighty Mandarin-speaking high-intermediate-level Japanese-as-a-foreign-language learners in Taiwan read a 1,336-word passage in L2 Japanese seven times and were instructed to read for meaning. The participants were also assigned to one of the four conditions: (1) listening + real-time/simultaneous oral reading (i.e., *shadowing*, which requires focal attention), (2) listening + oral reading at the end of each sentence (i.e., *time-lapse imitation*), (3) listening + silent reading (i.e., *subvocalization*, which requires the least investment of attentional resources), and (4) silent reading (i.e., *visual only*, which serves as single modality baseline; in addition to silent reading, the participants

were asked to utter a nonword sound [da] in their mouths every three to four seconds so as to prevent them from engaging in any read-alouds or subvocalization). On the following day, 21 multiple choice questions were given to the participants to measure their global, local, and higher order (e.g., inference) reading comprehension, and 36 items on the vocabulary assessment task to measure the target words. The vocabulary items in the post-test were the same as those tested in the pre-test two weeks prior to the treatment session.

After a one-hour treatment, the participants demonstrated significant gains for their receptive, form-meaning mappings of novel words on the vocabulary post-test. The participants had a mean score of 28.75% across the three conditions; i.e., approximately one out of three novel words. The mean score for the reading test was 82.29% (or 17.28 out of 21 correctness). A comparison of the four conditions revealed different patterns for the vocabulary measure (i.e., *shadowing > time-lapse imitation > subvocalization > visual-only* condition) and the reading measure (i.e., *subvocalization > time-lapse imitation > shadowing > visual-only* condition). The scores for the dual modality groups were all higher than the scores for the single modality (*visual only*) group. The researchers attributed the contribution of dual modality conditions outperforming the *visual only* condition for L2 reading comprehension to the use of phonological code, together with orthographic code, while making sense of printed information. However, significant differences were found only between *shadowing* (which requires focal attention) and *visual only* conditions for vocabulary test scores and between *subvocalization* or *time-lapse imitation* (which requires less investment of attentional resources) and *visual only* conditions for reading comprehension test scores. An

interesting result was found in that there was a positive correlation between the reading and vocabulary test scores when the reading test scores were equal to or above 15 points (i.e., comprehension of 71% and above for the reading text), and a negative correlation between the reading and vocabulary test scores when the reading test scores were below 15 points (i.e., less than 71% comprehension of the text).

Liu and Todd (2014) suggest that in dual modality repeated reading, directing learners' focal attention to phonological input while doing repeated reading (e.g., as the case of the *shadowing* condition) results in better vocabulary performance, while this same attention investment debilitates L2 reading comprehension. Constant focused attention to the formal features of the input would result in a less-favorable condition for meaning-based processing as a result of competing for limited capacity resource, but at the same time this would reverse the meaning-processing preference to reading and reorient the learners' attention resources toward form-based processing (Liu & Todd, 2014). On the other hand, tasks that require less investment of attention resources (e.g., *subvocalization*) to the phonological aspect of the repeated text are more beneficial to enhancing reading comprehension. As the default for reading is set for reading for meaning and the learners are successively revisiting the text through RR, less attention boost is required for comprehension-based or meaning-oriented reading (Liu & Todd, 2014). Liu and Todd concluded that the different amount of attentional investment required (for different forms of phonological support), the modality of learning preference (visual or auditory), and the purpose of the reading instruction (comprehension or acquisition), played a determining factor on the efficacy of dual modality input in repeated reading.

Nevertheless, several limitations of the Liu and Todd (2014) study should be noted. A limitation is that the effect of the repeated reading may have been mitigated for the silent reading, single modality baseline (*visual only*) group by requesting the participants to utter a non-word sound [da] every three to four seconds. While the researchers attempted to prevent the participants from engaging in any read-alouds or subvocalization, the participants' attention may have been distracted from processing the text. This could have led to a lower score for the *visual only* group and a larger difference when compared to the dual modality conditions. Had the *visual only* group participants not been distracted by having to utter a non-word sound, the results for the *visual only* group may have been higher and the differences between the dual modality and single modality conditions may have lessened or become non-significant. In lieu of the original design for the *visual only* single modality baseline group (silent reading + no subvocalization), subvocalization should be allowed alongside silent reading – as is the case for natural silent reading – to allow for comparison of the conditions.

Additionally, although Liu and Todd (2014) gauged both reading comprehension and vocabulary acquisition, a post-test was only provided for the vocabulary measure but not for the comprehension assessment. Even though the study incorporated the use of a comprehension measure, it can only document the participants' comprehension level after re-reading the text for other analyses (e.g., vocabulary acquisition and input modality comparison and correlation analysis), and the comprehension scores can only be used to conduct a post-treatment comparison between the groups. Without the use of a pre-test, the study is unable to document the participants' comprehension level prior to the treatment, to establish whether the groups' comprehension prior to the treatment was

comparable, and to document the changes, if any, in the participants' comprehension as a result of the RR treatment. Even though the study used random sampling, pre-treatment assessment is needed to systematically assess and assure comparability, or lack thereof, of the groups prior to the treatment. Therefore, Liu and Todd's (2014) claim that the dual modality groups significantly outperformed the single modality group on the comprehension measure is weakened by their study's limitations. Additionally, the treatment lasted only one session (a total of 1 hour), which does not allow the study to document the developmental changes across a study period (e.g., first, mid, last session) as a result of the RR treatment for both comprehension and vocabulary acquisition.

In a more recent study, Liu and Todd (2016) measured 80 Mandarin-speaking high-intermediate-level Japanese-as-a-foreign-language learners' incidental vocabulary gains through assisted RR (i.e., the same text read seven times) or assisted reading (i.e., seven thematically-related passages each read once). For the assisted RR condition, the participants were engaged in repeated encounters of the target words in the same text, whereas the participants in the assisted reading condition were engaged in repeated encounters of the target words in different texts. Identical to Liu and Todd's 2014 study described above, the participants were assigned to one of the four conditions: (1) listening + real-time/simultaneous oral reading (i.e., *shadowing*), (2) listening + oral reading at the end of each sentence (i.e., *time-lapse imitation*), (3) listening + silent reading (i.e., *subvocalization*), and (4) silent reading (i.e., *null, visual only*; including an articulation suppression technique). Participants were asked to provide an oral summary of the passage. The purpose of including such comprehension measure was only to ensure that the participants were focused on reading the passage for meaning; therefore, as the

researchers noted, the comprehension result was not the focus of further analysis. After a one-hour treatment, participants from both groups demonstrated receptive, form-meaning knowledge of one out of 3.44 novel words, with the assisted reading group (i.e., repeated encounters of the target words in different texts) showing higher gains. As Liu and Todd (2016) explained, through repetition of the unknown words in context, learners are provided with multiple exposure to "micro-level lexical constraints such a typeface, collocation, and orthographic features, as well as macro-level constraints such as contextual cues. Learners' episodic memory of the micro- and macro-constraints of novel forms from repeated reading sessions makes consequent recognition easier and faster" (p. 54).

To recap, researchers and practitioners in the recent decades have acknowledged the importance of facilitating L2 and foreign-language vocabulary acquisition through reading. Empirical research drawing on the efficacy of RR in L2 vocabulary acquisition has nonetheless been sparse, which underscores the importance for further investigation. The studies reviewed in this section seem to indicate that RR provided with external assistance led to vocabulary acquisition, but it remains uncertain whether such encouraging results hold if the learners are engaged in RR, with no or very little external assistance.

2.3.4 Repetition and Use of an Audio Model

Repeated reading research in the L2 has provided positive findings on the participants' perceptions of the effects of two features specific to RR: repetition in reading and audio model usage while reading (Taguchi, Gorsuch, Lems, & Rosszell, 2016).

With regard to repetition in reading, the participants in both Taguchi (1997) and Taguchi et al. (2004) have commented that the use of repetition in the RR treatment helped their reading fluency and comprehension. That is, repetition enabled readers to free their attentional resources and engage in higher-level comprehension monitoring. Studies conducted by Taguchi and Gorsuch and colleagues have reported that learners tend to gradually become aware of the benefits of repetition such as becoming more able to capture the details and development of a passage (Taguchi & Gorsuch, 2012).

While reading research has generally posed a positive effect and perception for repeated exposure to texts, some research (e.g., Taguchi & Gorsuch, 2012) raised the consideration regarding the optimal number of repetitions for a passage for each RR treatment. Taguchi and Gorsuch (2012) suggested that the number of repetitions may have an effect on students' motivation (to read) and an effect on the outcome. In considering the optimal number of repetitions for each passage, Taguchi and Gorsuch recommended striking a balance between the number of text repetitions per session and the effect of re-reading on learners' motivation.

Repeated reading research in the L1 English settings (e.g., O'Shea, Sindelar, & O'Shea, 1985; Samuels, 2006) reported that three to five re-readings are effective for an increase in reading rate and comprehension development. However, unlike L1 research which mainly dealt with younger participants, who seem to be able to engage in repeated tasks such as listening to audiobooks multiple times and still show enjoyment, adult readers generally find it boring and become impatient more easily if asked to perform the same task over and over again within a short period of time.

Previous RR-L2 studies (such as Taguchi & Gorsuch, 2002, 2012; Taguchi et al., 2004) have pointed out that "excessive repetitions (six or more) may discourage learners from fully engaging in rereading texts" (Taguchi & Gorsuch, 2012, p. 271). While more research exploring the optimal number of repetitions that would lead to fluency and comprehension development while preserving motivation is needed, the researchers recommended that three to five repetitions as suggested by precedent L1 studies are recommended for RR-L2 studies as a point of departure. In addition to learner motivation, factors such as learners' reading and language proficiency level and text difficulty may also affect the optimal number of repetitions and will need to be taken into consideration. Based on the findings from previous RR research in the L1 and L2, five repetitions were decided on for the current study to ensure a minimum threshold for comprehension and vocabulary acquisition to start happening while maintaining the reader's motivation.

With respect to the use of an audio model, RR research in the L2 has also provided positive perceptions of its usage and benefit in reading (Taguchi et al., 2016). Through the use of an audio model, the learner is presented with both the phonic and graphic forms of the reading passage. According to the dual modality theory, separate mental representations are activated for visual and auditory information (Baddeley, 2003; Lee & Young, 1974). The mental image provided by the written input and the phonological association provided by the auditory input complement each other (Anderson & Bower, 2013), and information presented simultaneously in the modalities of visual and aural input can enhance depth of information encoding, create more sophisticated memory traces, and facilitate retrieval and maximize retention of newly learnt information (Baddeley, 2003; Lee & Young, 1974). In terms of language learning,

that the readers are engaged with two channels of perception is suggested to (1) speed up

learning and improve the functioning of short-term memory, as well as (2) increase the

number of associations with a given word and increase the frequency and amount of

contact, which will result in better word retention and retention of grammatical

constructions in long-term memory (Dlugosz, 2000).

Based on the results of the extant RR-L2 studies, participants shared positive

comments on the usage of an audio model (i.e., assisted RR) (e.g., Gorsuch & Taguchi,

2008, 2010; Taguchi, 1997), which may have contributed to an increase in silent reading

rates and comprehension by providing additional prosodic information. The use of an

audio model facilitated the success of repeated readings and, as the learners in Taguchi's

(1997) study reflected, maintained their interest in the reading activity. Likewise, the

participants in Taguchi et al. (2004) reported that the use of an audio model improved

their textual comprehension, comprehension of new words, and motivation. In addition,

the RR group participants reported that assisted RR improved their listening skills; in

particular, their abilities to pronounce new words, which is an important lower-level

reading skill that many L2 or foreign language learners find difficult to develop (Gorsuch

& Taguchi, 2008, 2010). In Taguchi et al.'s (2012) study, Naomi verbalized that its utility

was beneficial in that the auditory component helped her pace her reading and enabled

faster reading, as well as assisted her text comprehension. The audio model also provided

her access to word pronunciation and prosodic information, which may have helped her

chunk the passage into meaningful clauses and thus facilitated comprehension. As

evidenced in L1 studies, the audio model provides appropriate phrasing of sentences for

the readers and prosodic features that assist readers in segmenting sentences into larger

and more meaningful units. This is beneficial for foreign-language learners who have relatively limited access to the prosody of the target language, and for those whose word recognition skills are not yet fully developed (Taguchi, 1997). Both the repetition and audio components of RR may have engendered motivation to read, as well as helped weak EFL readers develop fluency and become independent readers by providing a distinct form of scaffolding (Taguchi et al., 2004, 2012).

Studies show a preference for the use of the audio model in promoting L2 vocabulary acquisition. For instance, in Zahar et al. (2001) and Webb and Chang (2012), the participants were asked to subvocalize the sound code for the visual input, and results showed 7%-13% retention of novel words. In Han and Chen (2010), the participant performed time-lapse imitation of the audio model and was able to retain the orthographic form for 55% of the unknown words. Similarly, Liu and Todd (2014, 2016) found that the groups provided with auditory input (dual modality input), with or without oral production practice, performed better than the group that were engaged in merely silent reading; even though they found differences between the effects for the two modalities, the studies have limitations. As Han and Chen (2010) argue, in listening to and reading the text, the readers' phonological loop and visuo-spatial sketchpad of working memory (Baddeley, 1986, 2003) are both activated, thereby increasing the possibility for accurate processing and retention of the word.

Taguchi et al. (2012), on the other hand, did not find the use of audio assistance to be beneficial in all cases for their L2 learner (see also Therrien, 2004, for a similar negative view with respect to L1 reading). Taguchi et al. (2012) reported that the use of the audio model "sometimes inhibited [the participant's] comprehension by allowing her

little time to process ambiguities in the text. Also, text that contained difficult vocabulary or grammatical ambiguities was not made clearer through use of the audio model, thus resulting in her frustration" (p. 48).

The studies reviewed in this section seem to indicate positive effects of repeated encounter with the same text and leave open the question of the optimal number of repetitions for a passage under different conditions. Furthermore, Liu and Todd (2014, 2016) and Webb and Chang (2012) are the only studies to date that, to the best of my knowledge, teased apart and directly measured dual modality and single modality conditions with respect to vocabulary acquisition. However, Liu and Todd's (2014, 2016) studies have limitations, and the participants in Webb and Chang (2012) were provided with external assistance in addition to assisted/unassisted RR. Thus, little is known regarding the effects of dual modality input and single modality input on learners' vocabulary acquisition. This calls for the need for more research on this topic so that we can come closer to a conclusion in terms of which modality input (single or dual) is more beneficial for L2 vocabulary acquisition through repeated reading.

2.3.5 Other Beneficial Effects

Empirical and pedagogical evidence in L1 research has shown that RR helps L1 readers gain confidence in their reading, motivates them to read more, and enhances self-monitoring, an essential behavior for independent readers (Koskinen & Blum, 1984). In RR-L2 studies, similar benefits have been found.

For example, in Taguchi et al. (2004), the participants' responses on the questionnaire suggested that both the RR and ER methods encouraged their development as independent readers: the participants commented, specifically, on an increase in their

willingness to read longer texts and their use of metacognitive reading strategies. Additionally, the researchers found that the RR and ER approaches enabled their participants to learn how to deal with unknown words by using contextual clues or skipping unimportant words, which is suggestive of longitudinal growth in the participants' ability to use various reading strategies. Furthermore, the methods provided them extended exposure to a substantial amount of L2 input through reading that may result in positive effects on their L2 development and vocabulary growth (Taguchi et al., 2004).

In Gorsuch and Taguchi's (2010) study, aside from the participants' perceptions of the changes in reading speed and comprehension, the researchers also reported four other categories that emerged from the participants' reports: (1) reading strategy use, (2) motivation to read, (3) language skill development, and (4) changes in attentional resource demands. The participants reported use of a variety of higher-level post-lexical reading strategies, including cognitive strategies (i.e., bottom-up and top-down) and metacognitive strategies (e.g., monitoring comprehension), which as Koda (2005) suggested are the "decisive metacognitive capability separating good from poor readers" (p. 212). Results further revealed a decrease in the use of bottom-up strategies, an increase in top-down strategy use, and no changes in the use of metacognitive strategies throughout the study period. In terms of motivation to read, the learners expressed an increase of motivation for reading, both short-term and long-term interests. The learners also commented on language skills development and perceived that assisted RR improved not only their reading skills but also their language skills, including listening, grammar, pronunciation (segmental and suprasegmental aspects), vocabulary, writing,

speaking, and discourse awareness. As Gorsuch and Taguchi (2010) discussed, the participants' improvement in word decoding and comprehension allowed them to allocate their attention elsewhere. Participants' reflections on their changes in attentional resource demands revealed a shift from pre-lexical processing, grammatical structures, pronunciation, and remembering details, to increasing use of post-lexical processes. On the whole, Gorsuch and Taguchi's (2010) qualitative, longitudinal study supported the quantitative findings of their 2008 study, which showed that assisted RR treatment has a positive effect on EFL learners' reading fluency and comprehension, in addition to contributing to their general language development and helping them become independent readers. Through extended experience with RR, the readers showed an increased use of metacognition in reading strategy and a heightened awareness of the relationship between fluency and comprehension (Gorsuch & Taguchi, 2008, 2010).

2.4 Summary

In summary, this review of the literature has discussed major theoretical concepts as well as empirical studies relevant to the repeated reading approach in the L1 and L2 settings. L1 research has provided converging evidence that RR is effective in facilitating L1 fluency and comprehension. L2 research has substantiated positive but mixed findings for the effects of the RR approach on reading fluency and comprehension and vocabulary acquisition. From the review of the literature, several gaps remain in research.

The literature indicates that in both the L1 and L2, repeated reading of the same text leads to improvement in reading comprehension. However, findings were less conclusive as to whether L2 repeated reading leads to a transfer of practice effect to new,

unpracticed passage in terms of reading comprehension. Also, there is limited research investigating the effects of L2 repeated reading on vocabulary acquisition, which underscores the importance for further investigation. Past research indicates that RR provided with external assistance leads to vocabulary acquisition, but it remains uncertain whether RR with no or very little external assistance will likewise lead to vocabulary acquisition. Therefore, the present study filled these gaps by investigating whether L2 repeated reading promotes reading comprehension of a new, unpracticed passage and leads to incidental vocabulary acquisition.

Additionally, while previous RR research findings in L1 documented the potential of RR for difficult texts (Hyönä & Niemi, 1990; Moyer, 1982), little is known regarding L2 readers' performance when provided with difficult text. In RR-L2 research, the majority of existing studies have utilized graded readers and level-appropriate texts in their studies. Han and Chen (2010) was the only study found that utilized difficult texts, and the heritage speaker participant (when provided with a variety of instructional strategies) demonstrated a significant amount of vocabulary acquisition and the ability to process texts that were far beyond her capacity to decode. Also, more of the existing RR-L2 studies provided the participants a variety of instructional strategies. Thus, more research is needed to see whether the participants would still benefit through repeated reading when they are provided with more difficult texts and no additional support. Furthermore, authentic input should be prioritized over modified input in that authentic input contains features required for acquisition to occur and is in line with the communicative approach which aims at fostering the learners' ability to work with authentic texts (Han & D'Angelo, 2009). Thus, the present study utilized an authentic

text (i.e., originally written and unmodified text) and a more challenging text to measure the effects of L2 RR when the participants are provided with authentic and challenging reading materials and no additional support.

Due to the limitations, difference in focuses, as well as methodological differences in the limited RR-L2 studies – e.g., assisted and/or unassisted RR, RR only vs. RR-based (form-focused instructions, or production exercises), frequency of treatment, and number of repetitions – researchers have not yet arrived at a conclusive view regarding the efficacy of various forms of the RR instruction. Additionally, many previous repeated reading (RR) studies in the L2 did not conduct a systematic study of the contributions made by different components of RR. For some of the RR-L2 studies (e.g., Chang, 2012; Gorsuch & Taguchi, 2008; Taguchi, 1997; Taguchi & Gorsuch, 2002; Taguchi et al., 2004; Taguchi et al., 2012), a single treatment session included both assisted RR and unassisted RR, and the two conditions were administered differently (e.g., procedure and time on task). Thus, it becomes hard to identify the extent to which the change was due to assisted or unassisted RR, let alone other instructional strategies that were used to accompany repeated reading. Studies should tease apart the effects of assisted or unassisted RR, among other variables, in their design. Furthermore, there were variations with the use of an audio model, with some studies asking their participant(s) to listen to the audio model and others requiring an additional phonological production practice (e.g., Han & Chen, 2010; Liu & Todd, 2014, 2016). As Liu and Todd (2014) argue, a systematic study of the different forms of phonological practice is warranted to garner a better understanding of the role and nature of the phonological component in RR. Therefore, studies should employ designs that controls for various components of RR and

isolates their contributions. To fill this gap, the present study sought to compare the effects of unassisted RR and assisted RR conditions.

The extant reading research studies have however established the following categorical findings worth noting for the present study: (1) The use of context is found to facilitate learners' word recognition skills and reading comprehension. Vocabulary is acquired most effectively "through frequent encounters with words in multiple contexts" (Han et al., 2009, p. 12); (2) Assisted reading or assisted RR seems to aid fluency, comprehension, and vocabulary acquisition (in particular, incidental vocabulary acquisition) more than unassisted reading/RR. For incidental vocabulary acquisition, researchers have found that dual modality input is more effective than providing merely one source of input, written or oral, to the learners (Brown, Waring, & Donkaewbua, 2008). Likewise, Nation (2001) argued that contextualized spoken input is an effective means of learning and building one's vocabulary knowledge; (3) Studies have documented positive effects of repetition in facilitating reading fluency and comprehension. They have also documented positive effects of frequency in successful vocabulary acquisition (Grabe, 2009; Nation, 2001; Zahar et al., 2001). According to the literature, the number of repetitions required for a learner to remember a word varies, strongly depending on various factors including learner differences (Tinkham, 1993), the amount of time spent on each repetition, and the depth of information processing (Craik & Lockhart, 1972).

2.5 Research Questions

Whereas previous studies explored the efficacy of repeated reading on either reading fluency in L1, fluency and comprehension of foreign-language readers, *or* vocabulary acquisition of a heritage speaker or foreign-language learners, the present study explores an under-investigated area in L1 and L2 research by examining how L2 repeated reading – which included the provision of an audio model, or without – simultaneously affects reading comprehension and incidental vocabulary acquisition for EFL learners. This study is guided by the following research questions:

Research question 1: Does L2 repeated reading promote reading comprehension?

Research question 2: Does L2 repeated reading lead to incidental vocabulary acquisition?

Reading comprehension refers to the ability to understand the reading text and is operationalized as the amount of text information recalled. Vocabulary acquisition is defined as knowledge of form-meaning mappings of novel words and operationalized as the ability to recognize (i.e., receptive knowledge) and/or produce (i.e., productive knowledge) words in context and isolation (Han & Chen, 2010). Incidental learning refers to the learning mode in which participants are not forewarned of an upcoming test measuring their acquisition and retention of a particular type of information (Hulstijn, 2005). In the case of incidental vocabulary acquisition, learners do not explicitly study (but may have inferred) the word meaning while reading, and vocabulary acquisition is a by-product of reading comprehension, rather than a focus in its own right (Han & Chen, 2010). In this study, receptive word knowledge refers to learners' ability to recognize the L2 word form and provide the L1 meaning of the target word, whereas productive word

knowledge requires learners to demonstrate their ability to produce the L2 word form from the L1 meaning provided.

The present study employed a mixed-methods design. A primary goal of the study is to examine and compare the participants' performance through quantitative data (i.e., pre-test, post-test, and delayed post-test scores). Immediate written recall protocols were used to gauge the learners' reading comprehension development, whereas vocabulary tests were implemented to measure incidental vocabulary acquisition and retention. A secondary objective is the analysis of qualitative measures. Specifically, qualitative analyses of the participants' learning journals, background questionnaire, and surveys, were included to yield additional insights about the effects of RR on L2 reading comprehension and vocabulary acquisition.

Chapter III

METHODOLOGY

This chapter describes the methodology employed in the present study. It will first describe methodological issues of previous studies and provide a brief summary of a pilot study, with a focus on the results and limitations that informed the methods used in the present study. Next, the design for the present study and a description of the participants are presented. This is followed by a presentation of the methods of data collection, including the materials, target words, measurement instruments, and procedures for data collection. The chapter ends with an overview of data coding and data analysis procedures.

3.1 Methodological Issues

The review of the literature revealed several methodological issues when researching reading and vocabulary acquisition, and the most notable ones are outlined as follows.

A first methodological issue concerns the lack of a control group. A few previous experimental studies (e.g., Chang, 2012; Taguchi, 1997) that compared the effect of repeated reading under various treatment conditions over extended period of time did not include a control group. Some studies claimed an inclusion of a control group, but the participants were required to perform additional tasks. For example, the participants were also given reading practice (such as extensive reading) (e.g., Taguchi & Gorsuch, 2002; Taguchi et al., 2004) or received intensive reading and grammar instruction in another

class (e.g., Gorsuch & Taguchi, 2008, 2010). The participants in Liu and Todd's (2014) control group were distracted by having to utter non-word sounds and not being allowed to engage in subvocalization when performing silent reading. Thus, the so-called control group in these studies was not a genuine control group, but rather a comparison group. This would most likely have affected the learners' performances and contributed to the lack of differences between the RR and the control groups shown in their studies. The results may have been different with a control group that was strictly controlled for enriched reading experience. Hence, the present study included a genuine control group to support that the improvement, if any, was a result of the treatment.

Another issue with the design of several previous studies is the repetitive use of the same comprehension questions within a very short period of time, which may have led to practice effect. For instance, in many of the RR-L2 studies, the same comprehension questions were given to the participants after the first, third, and seventh readings (e.g., Taguchi & Gorsuch, 2002), after the first and fifth (final) readings (e.g., Chang & Millett, 2013; Gorsuch & Taguchi, 2008), or after almost every reading (e.g., Taguchi et al., 2012). This may have affected the participants' understanding of the text and/or reading strategies they employed and resulted in an increase in their comprehension scores that was actually a result of practice effects or revisiting the same test content within such a short period of time. To control for practice effects found in many previous studies, the present study administered comprehension tests only after the fifth (final) reading.

A third issue involves a discrepancy with regard to the reading passages or the testing questions. Firstly, a main limitation concerns a lack of similarity of the readability

level and of the topic between the treatment and the testing passages, and between the pre-test and the post-test passages (e.g., Taguchi & Gorsuch 2002; Taguchi et al., 2004). The studies' limitations may have caused the results to provide only modest support for the RR approach. Such incomparability between the passages may have contributed to the lack of differences between the RR and ER groups, and the lack of clear transfer effects, for both reading rate and comprehension. In Chang and Millett's (2013) study, while the testing passages were comparable in terms of vocabulary levels, the unpracticed passage contained easier grammatical structure, more complicated plot, and seven times more words, than the practiced passages. The passages being non-equivalent in difficulty and length makes it challenging to determine and compare the treatment effects. As reported in the literature, such similarities in terms of vocabulary overlap and post-lexical factors such as schemata and syntactic difficulty are critical factors in the success of the RR treatment (Taguchi & Gorsuch, 2002). Thus, by utilizing a whole story that was written by the same author and has the same story plot, the writing style, voice/tone and topic for all the passages can be controlled for. Researchers have pointed out that a prerequisite for a high transfer of gains in reading rate to the new passage lies in a high degree of overlapping words between the old and new passages, which can be maximized by using a whole story or a series of related passages on the same topic (Rashotte & Torgesen, 1985). Thus, the present study attempted to control for difficulty of the passages and utilized a whole story as reading material.

Additionally, there was a discrepancy in the difficulty level and a lack of similarity between the comprehension questions in the pre-test and the post-test (e.g., Gorsuch & Taguchi, 2008). Also, the questions may have lacked validity or turned out to

be exclusively detail oriented, resulting in the participants focusing only on information needed to answer the questions in subsequent re-readings (e.g., as is the case for the short answer test designed to capture main ideas, supporting ideas, and details in Gorsuch and Taguchi's 2008 study). They also might not have been sensitive enough to capture the participants' comprehension gains (e.g., Taguchi & Gorsuch, 2002; Taguchi et al., 2004). Thus, the present study attempted to control for these issues.

A fourth issue involves the shortness of the treatment and a lack of delayed post-tests. For instance, for a few of the studies (e.g., Liu & Todd, 2014, 2016), the treatment lasted only one session and a total of an hour. This does not allow for the documentation of the effects of the RR treatment for comprehension and vocabulary development over time. Also, short treatments do not provide a sufficient amount of time for incidental learning to occur. On a similar note, it is worth mentioning that none of the aforementioned RR-L2 studies on vocabulary acquisition incorporated a delayed post-test. However, only with the usage of a delayed post-test can we document whether any gains were retained several months later. For this reason, the present study incorporated multiple treatment sessions and a delayed post-test.

A final issue concerns the limitations of commonly used comprehension measures. The present study attempted to overcome the limitations by using the recall protocol. The ensuing paragraphs will first introduce the recall protocol and discuss how the various limitations of the commonly used comprehension measures can be overcome by the recall protocol. This will be followed by a consideration of the challenges of the recall task and the means the present study employed to overcome these limitations.

The written recall task has been widely used by L1 literacy and psycholinguistic researchers (e.g., Fitzgerald & Spiegel, 1983) and in L2 reading research (e.g., Bernhardt, 1983a, 1983b; Carrell, 1983), and it has been shown to be a valid means of measuring reading and listening comprehension (Bernhardt, 2011; Chang, 2006; Lee, 1986; Riley & Lee, 1996). It requires the participants to read or listen to a text, of preferably 200-250 words, and then write out as much as they can remember from the text without referring back to it. Even though free written recall is not an exact measure of online comprehension, recall protocols provide a general indication of comprehension (VanPatten, 1990). The underlying assumption is that the recall protocols reflect the relative degree of attention that the learner is able to pay to the content and that a lower recall score may indicate a disruption in attention to content and, thus, in comprehension (VanPatten, 1990). As the information recalled is learner-generated, the written recall task, in Johnston's (1983) words, is "the most straightforward assessment of the result of the text-reader interaction" (p. 54). It is considered a direct, construct valid, and integrative assessment task of what the learner has understood after reading a passage (Berkemeyer, 1989; Bernhardt, 1983a, 1992; Lee, 1986).

Johnston and Bernhardt (as cited in Berkemeyer, 1989), among many other researchers, have highly advocated the virtues of immediate written recall protocols in evaluating reading comprehension. First, unlike traditional tests such as multiple-choice questions that may invite guessing (Thompson, 1995), learners are unable to guess their way through a recall task. Second and more importantly, the recall task allows misunderstanding or gaps to emerge, which is a virtue that other evaluation procedures cannot offer (Berkemeyer, 1989). The recall protocol as a testing and/or diagnostic

procedure does not test for grammar points and directs the readers' attention to the meaning and meaning construction of a passage (Wells, 1986). Third, unlike traditional tests (e.g., multiple-choice tasks and short answer questions) that may delimit possible answers and give away additional hints or information, generating recall data does not interfere with the learners' comprehension of a text (Bernhardt, 1991; Chang, 2006; Liao, 2009).

Despite the main advantages presented above, potential challenges of using the recall protocol in assessing comprehension include difficulties in controlling for memory factor, difficulties in scoring recalls, and problems imposed by one's writing ability (Alderson, 2000; Liao, 2009). The first challenge of using the recall protocol to assess reading comprehension lies in difficulties in controlling for memory factor. Research (e.g., VanPatten, 1990, and successive replication studies) has typically yielded relatively low number of recalls (approximately 33% and less), which raises a potential reliability problem of the written recall task that may be undermined by the role of individual memory capacity (Leow, Hsieh, & Moreno, 2008). The total number of words in the passages for VanPatten's study (1990) and successive replication studies was 274 words. A way to decrease the potential threat caused by this problem would be to decrease the length of a passage. Thus, the present study limited the number of words to 200-250 per passage. Additionally, in van Zeeland and Schmitt's (2013) study, they asked the participants to listen to the passage twice with the purpose of avoiding memory effects on the comprehension results. This implies that revisiting the text through multiple readings may help decrease the effect of the memory factor on the comprehension results (Schmitt et al., 2017). The participants in the current study were asked to read the passage five

times. Through controlling for the length of a passage and repeated exposure to the text, the potential threat caused by the memory factor should be minimized.

A second disadvantage of the recall protocol centers on efficiency of scoring the protocols in terms of the amount of time required for developing a scoring scheme as well as the number of hours involved in scoring itself (Bernhardt, 1992). More importantly, the recall protocol has been criticized for a lack of a valid scoring method (Wells, 1986). The present study concurs that a construct valid scoring procedure that reflects the idea units' relative importance to the overall meaning of the passage should be employed for coding. As idea units do not carry the same weight or value (Wells, 1986), scoring rubrics should thus be hierarchically weighted to reflect their relative importance for a passage. Past studies (e.g., Carrell, 1985) have generally used a three-level coding scheme for their expository texts: High-level (i.e., main idea), Mid-level (i.e., supporting idea), and Low-level (i.e., details) idea units. In the present study, a pilot coding on a narrative text has found that the three-level scoring scheme may not be as suitable for the narrative text – as reflected by a high inter-coder reliability for the main-idea idea units ($r=.97$) but low reliability in terms of the mid-level (supporting ideas) and low-level (details) idea units ($r=.43$). This makes logical sense in that expository texts tend to have clearer hierarchical structures and can easily be hierarchically structured into three levels. Pilot coding, however, have suggested that for narratives, the level distinctions were not as clear-cut in terms of the "supporting ideas (i.e., Mid-level)" and "details (Low-level)" categorization in Carrell's scoring scheme. Therefore, the present study adapted existing scoring schemes into a two-level hierarchy to reflect "Main-idea" and "Detail" information of the narrative passage, as well as assigned different scores to

the idea units to achieve construct validity. (See the Data Coding section for a detailed description on the scoring scheme.)

A third disadvantage of the recall measurement is the possible disparity between the participants' comprehension and production abilities (Lund, 1991). As recall tasks involve written protocols, the participants' level of L2 writing may not reflect their level of reading comprehension, which could impose a potential threat on the validity of the measure (Liao, 2009). Previous L2 research that examined comprehension assessment tasks utilizing written recall protocols has identified language of recall as a key variable for learners at the beginning and intermediate levels, but not at the advanced levels (Brantmeier, 2006; Lee, 1986). For participants with lower proficiency levels in particular, recalls should be completed in the participants' native language so that it is a test of L2 reading/comprehension rather than a test of L2 writing (Alderson, 2000) and that the participants' comprehension are not masked by their limited production abilities in the second language (Chang, 2006; Lee, 1986). Thus, in the present study, the intermediate-level participants were asked to recall in their L1.

3.2 Pilot Studies

An initial pilot study was conducted on two classes of English-as-a-foreign-language (EFL) learners who were from the same grade level, major, and vocational high school as the participants in the present study. The purpose of this pilot study was to verify goodness of fit of design and procedures with respect to learner proficiency level and learning context. The results were used to inform decision-making and modification to the current study design.

In this pilot study, 70 students were randomly assigned to either the Assisted Repeated Reading group (i.e., repeated reading along with an audio model) or the Unassisted Repeated Reading group (i.e., repeated reading only). A week prior to the treatment, the participants completed an extensive background questionnaire, took a vocabulary pre-test, and identified unknown words in the reading passage. During the treatment phase, the participants read the same passage seven times. Then they completed a comprehension post-test and a vocabulary post-test, followed by a survey.

A 477-word segment from E. B. White's *The Trumpet of the Swan* was used for both treatment and testing purposes. The passage had a VocabProfile of 81.86% at the 1-1000 word level, 6.19% at the 1001-2000 word level, 0.62% AWL Words, and 11.34% Off-List Words. According to Hu and Nation (2000), the reader would need to have a vocabulary size at the first 1000 level in order to comprehend 84.8% of novels for younger learners, and an additional vocabulary size at the second 1000 level to comprehend 90.7% of novels for younger learners.

Table 1

Pilot Study Known Word Density

Known word density	Groups		Total (percentage)
	Unassisted RR (n=35)	Assisted RR (n=35)	
98%-100%	9	7	16 (22.86%)
95%-98%	6	13	19 (27.14%)
90%-95%	18	12	30 (42.86%)
< 90%	2	3	5 (7.14%)

Out of the 70 students piloted, 16 students (22.86%) had a known word density of 98%-100%, 19 students (27.14%) had 95%-98%, 39 students (42.86%) had 90%-95%, and five students (7.14%) had a known word density of <90% (see Table 1). Given that

most of the participants seemed to demonstrate 90% or above text coverage, this suggests

that most of the students from this population would have an estimated vocabulary size at

the second 1000 level. Results indicated that the majority of participants for the current

study would have the ability to read texts written within the 2000 high frequency words.

There were five lessons learned from the pilot study that had implications for the

present study. The first concerns the number of reading times. Based on the responses

from the learning journals and exit questionnaire, many participants reported that three to

five re-readings were sufficient to understand the text. Additionally, many of them felt

that it was too tiring and redundant to have to read the text seven times. The students

reported that they were able to focus their attention up to the fifth re-reading, after which

their attention waned. This is consistent with the literature (e.g., Chang, 2012) where the

participants reported feeling tired after the fourth or fifth re-reading. Therefore, the

number of reading times was adjusted to five re-readings for the present study, in order to

allow for learning to take place and minimize fatigue. A second lesson concerns the

length of each passage. A limitation of the pilot study was that the 477-word segment was

too long a text length for the participants to recall, as reflected in their feedback and test

results. Also, Dowhower (1989) suggested that shorter passages of 50 to 300 words are

best suited for repeated reading. Therefore, the length for each segment was kept to 200-

250 words per passage for the present study. Third, based on the results and feedback, the

amount of time allowed for each task was adjusted for the current study, as will be

described in the Procedure section. Fourth, the participants in the present study were to

recall in their L1, in lieu of the L2, so that the recall task is a test of reading

comprehension rather than a test of L2 writing (Alderson, 2000); researchers have

reported that the language of recall plays an influential role on the participants' performance, especially for intermediate-level learners (Brantmeier, 2006; Lee, 1986). And lastly, the participants did not receive any EFL instruction from the teacher or researcher so that the external assistance would not become an extra variable for the study.

The rationale behind the selection of reading material was based on several considerations: lessons learned from the pilot study and the literature (e.g., length of passage per session, the use of a whole story that was written by the same author and has the same underlying story, and the participants' proficiency level), the participants' interests, and the number of sessions that the participants were able to partake. Prior to the current study, an informal survey was conducted with the participants, which consisted of mostly adolescent male participants, to determine their preference of the genres of novels. The participants' interests were taken into account to enhance learner motivation. The results indicated that they favored narratives with adventure or mystery or action. Also, logistic constraints were taken into consideration for the length of the entire story so that the number of available sessions (i.e., 11 reading sessions) matched with the length of each passage, when the story was segmented into 200-250 words per passage. Taking all the criteria into consideration, *The Tale of Samuel Whiskers* was selected as reading material for the current study.

Prior to the actual data collection, the researcher re-piloted the new procedures and materials, using the same reading passages and tasks as the current study, on five students who did not participate in the initial pilot study and were also from the same grade level and school. A main purpose of this final piloting was to ensure the flow of the

experiment and to double check whether any minor modification was needed. The results indicated that the story was challenging enough and interesting to read, and the number of words per session was workable.

3.3 Current Study

3.3.1 Design

The present study employed a pre-test, treatment, post-test, and delayed-post-test design. As shown in Table 2 below, the participants attended 17 sessions, eight of which were for treatment purposes (Sessions 4-11) and nine for testing purposes (Sessions 1-3 for the pre-test, Sessions 12-14 for the post-test, and Sessions 15-17 for the delayed post-test), over a period of four months. The participants were assigned to one of the three conditions: Unassisted RR, Assisted RR, or Control group.

The entire study took place in a computer lab, and the participants were monitored by the researcher, along with their advisor and/or a research assistant. The materials were presented to the participants, and data were collected, through Qualtrics, which is an Internet-based data collection and analyzing platform for academic and business research. The participants were not aware of the actual purpose of the study. Constant consultation with their English teacher and advisor took place throughout the study.

A flow chart of the background preparation procedure is illustrated in Appendix A for the reader's reference.

3.3.2 Participants

The participants initially consisted of 122 lower-intermediate-level EFL learners (16-17 years old) from three intact classes in the second year of a vocational senior high

school (eleventh grade) in northern Taiwan. The participants from each of the three classes were assigned as a group to one of the three conditions: Unassisted RR (N=41; repeated reading only), Assisted RR (N=42; repeated reading while listening to the audio model), or Control (N=39) group.

Table 2

Study Design

	Unassisted RR group (N=41)	Assisted RR group (N=42)	Control group (N=39)
Preparation	(1) Informed consent (2) Online background questionnaire (3) English proficiency and learning information gathering		
Pre-test (Sessions 1-3) [Week 1]	(1) Vocabulary pre-test (2) Identification of unknown words (3) Repeated reading (pre-test passage) (4) Comprehension pre-test (5) Self-evaluation form		
Treatment (Sessions 4-11) [Weeks 2-5]	(1) Identification of unknown words (2) Repeated reading *w/o audio model* (3) Learning journal (4) Self-evaluation form	(1) Identification of unknown words (2) Repeated reading *w/ audio model* (3) Learning journal (4) Self-evaluation form	
Post-test (Sessions 12-14) [Week 6]	(1) Vocabulary post-test (2) Identification of unknown words (3) Repeated reading (post-test passage) (4) Comprehension post-test (5) Self-evaluation form (6) Exit survey		
Delayed post-test (Sessions 15-17) [Week 16]	(1) Vocabulary delayed post-test (2) Identification of unknown words (3) Repeated reading (delayed post-test passage) (4) Comprehension delayed post-test (5) Self-evaluation form (6) Exit survey		

The participants were native speakers of Chinese and learned English in a foreign language setting. The EFL setting was chosen to reduce the possibility of the participants'

performance (e.g., test scores) being inadvertently affected by exposure to a large amount of L2 input that is characteristic of ESL settings. Learners from this population had received similar compulsory elementary and junior high school education (first to ninth grade), thereby making the English education consistent for all participants. By the time of the study, all of the students completed at least 6.5 years of formal EFL education. By the time Taiwanese students graduate from junior high school, they are expected to have mastered at least the 1200 word list announced by the Ministry of Education, because the EFL textbooks and curricula are required to cover all the words on the list. Even for low achievers, a minimum of 448 words from the list is required to graduate from junior high school. By the time students in Taiwan graduate from senior high school, they should have knowledge of 2000 general service words, which is the high school graduation threshold (Hsu, 2014). Consultations with the participants' current English instructor revealed that their current EFL education was limited to approximately 4 hours of instruction per week and the prescribed content: that is, textbooks and formal classroom instruction. An analysis of the participants' EFL textbooks, instructional activities, homework assignments and midterm/final exam papers revealed that for reading comprehension tasks, there was a heavy reliance on multiple choice, cloze test, fill-in-the-blank (idioms), sentence translation, and sentence combination or reordering (with a focus on testing grammar and vocabulary).

Detailed information about the participants, including their English learning condition, was collected via the online background questionnaire that was administered through Qualtrics (Appendix B; See also the Instruments section for a detailed description on the questionnaire). The participants reported that they had limited contact

with the English language outside of school instruction (in the past six months), as a majority of them had either been fully occupied with school exams (i.e., daily, midterm/final) and other course assignments, or were afraid of the English subject and lacked internal motivation to learn the target language. Their major at the vocational high school was Control Techniques, which aims to provide the students control technical skills in the electrical technique profession, and the majority of the school curriculum centers on engineering, automatic control, mathematics, chemistry, and practical training. Thus, their primary contact with the English language constituted merely materials required by the curriculum.

An initial round of analysis was administered to eliminate the participants that did not meet the pre-set criteria. Detailed information regarding the elimination criteria will be presented in the Data Analysis section. Additionally, the participants were checked for outliers. Results from following numbers of participants are presented and discussed in the following chapters: Unassisted RR (n=14), Assisted RR (n=15), and Control (n=13) group.

Lastly, comparability of the groups was checked to ensure that there was baseline equivalence of the groups. Additional information about the participants' English proficiency and learning were collected, including the their six English midterm and final exams scores from the whole academic year. An analysis of the participants' English school exam scores was conducted on the remaining participants (i.e., after the eliminating process) to check whether the English proficiency of the participants from the three groups was comparable. A one-way ANOVA was run on the participants' mean scores (Unassisted RR group: *M=66.07, SD=15.91*; Assisted RR group: *M=68.09,*

SD=14.40; Control group: *M=67.06, SD=9.97*), and the results indicated that there was no statistically significant difference between the groups (*F(2,40) = .082, p = .921*). In addition, the scores were checked for baseline equivalence of the groups by checking if the group mean differences were less than or equal to one-quarter of the pooled standard deviation, which indicates that the groups are equivalent (see the Data Analysis section for a detailed description on the baseline equivalency check) (Ho, Imai, King, & Stuart, 2007; National Center for Education Evaluation and Regional Assistance, 2014). Results indicated that baseline equivalence of the groups was met for the participants, as the group mean differences (Assisted RR – Unassisted RR = 2.03; Control – Unassisted RR = 1.00; Assisted RR – Control = 1.03) were less than one-quarter of the pooled standard deviation (i.e., 3.43).

3.3.3 Data Collection

3.3.3.1 Materials. *The Tale of Samuel Whiskers*,[1] written by Beatrix Potter, was selected as reading material (see Appendix C). The 2616-word extended narrative was segmented into 11 consecutive passages. Among the 11 passages (*Mean* = 235.54; *Range* = 201-251 words), Passage 1 was used for the pre-test, Passages 2-9 for the eight treatment sessions, Passage 10 for the post-test, and Passage 11 for the delayed post-test. By the time the two treatment groups participants completed the entire study, they had

---

[1] The tale is a story of Tom Kitten who lives with his cat mother, Mrs. Tabitha Twitchit, and two sisters in a house overrun with rats. It illustrates what happens when Tom Kitten escapes his mother's control and accidentally comes across the rat Samuel Whiskers' secret hideout behind the attic walls, how Samuel Whiskers and his wife Anna Maria attempt to make Tom Kitten into a pudding and then eat him, how Tabitha Twitchit and her cousin in search for Tom Kitten call for the help of a carpenter (John Joiner) to saw open the attic floor and rescue Tom Kitten, how Samuel Whiskers and Anna Maria flee the attic and make their escape to Farmer Potatoes' barn nearby, and lastly, how this incident causes Tom Kitten to become afraid of anything that is larger than a mouse.

finished the entire story; the control group participants only read the three testing passages. Audio recordings of a native English speaker reading the passages at a natural pace were provided as an audio model for the Assisted RR group, and headphones were provided for each student during the reading process.

3.3.3.1.1 Analysis of text. To measure difficulty of the text and ensure comparability of the reading materials, the passages were analyzed using the VocabProfile-Compleat from LexTutor (Cobb, 2016). As shown in Table 3, the numbers of words in the three testing passages are similar (i.e., 245, 251, and 237 words, respectively, for the pre-test, post-test, and delayed post-test passages), and the eight treatment passages have a mean word count of 232.25 (range: 201-249 words). In terms of type-token ratio, the testing passages again show a similar type-token ratio for the pre-test (0.59), post-test (0.57), and delayed post-test (0.59) passages, suggesting that the testing passages' type-token ratio are comparable. The type-token ratio for the treatment passages, however, is lower (0.31) than the testing passages, indicating less lexical variety and greater vocabulary overlap for the treatment passages. The type-token ratio for the entire story is 0.28 (747 types and 2638 tokens). The percentage of words belonging to the 1-K + 2-K word levels groups for each passage is also similar across the passages, ranging from 83.33% to 85.66%. According to these indices, the results suggested that each of the treatment and testing passages represented a similar level of reading difficulty and that the segments are comparable in terms of the vocabulary level of and passage length.

A post-hoc text analysis was conducted for the three testing passages using English Lexile Analyzer (www.lexile.com), which yielded results of the Lexile Measure

and the Mean Log Word Frequency. The Lexile Measure refers to the reading demand of the text in terms of the semantic difficulty and syntactic complexity, and analysis showed that the Lexile Measure for the pre-test, post-test, and delayed post-test passages were 920L, 1070L and 800L, respectively. The Log Word Frequency refers to the logarithm of the number of times a word appears in each 5-million words of the MetaMetrics research corpus of 650 million words, and the Mean Log Word Frequency refers to the average of all such value for all of the words in the passage. Analysis revealed that the Mean Log Word Frequency was 3.66, 3.73, and 3.72, respectively, for the three passages.

Table 3

Text Analysis

	Pre-test (Passage 1)	**Treatment (Passages 2-9)**	**Post-test (Passage 10)**	**Delayed post-test (Passage 11)**	**ALL passages (Passages 1-11)**
Number of words	245	1883 (*Mean: 232.25, Range: 201-249*)	251	237	2616 (*Mean: 235.54, Range: 201-251*)
Type-token ratio	0.59	0.31	0.57	0.59	0.28 (747 types; 2638 tokens)
K-1 words K-2 words (cumul)	77.96% 6.12% (84.08%)	77.06% 6.27% (83.33%)	80.08% 4.38% (84.46%)	77.22% 8.44% (85.66%)	77.45% 6.27% (83.72%)
Lexile Measure	920L		1070L	800L	
Mean Log Word Frequency	3.66		3.73	3.72	

It is worth noting that the post-test passage seemed to be slightly more difficult than the pre-test passage and the delayed post-test passage with respect to the reading demand of the text in terms of the semantic difficulty and syntactic complexity. To achieve comparability of the testing passages, the study controlled for factors such as:

similar numbers of words (word count) and type-token ratio, percentage of high frequency words (i.e., K-1 and K-2 level words), writing style and topic (i.e., the same author and the same story plot), and Mean Log Word Frequency for the three testing passages. Despite that, the Lexile Measure was still found to be different for the three passages.

     3.3.3.1.2 Participants' known word density. In order to determine the participants' known word density, all groups were given the pre-reading task that entailed identifying all unknown words for each of the passages they read. The participants were provided with sentences that were broken down from the passage, and the order of the sentences was scrambled. The sentences also consisted of non-words, which were inserted and evenly distributed in each sentence. This was to prevent the participants from overstating their vocabulary knowledge by indicating unknown words as known (Anderson & Freebody, 1983; see, also, Zimmerman, Broder, Shaughnessy, & Underwood, 1977, for description of the "lexical decision task") and to check whether the participants were careful in performing the task: that is, whether the participant did not identify a non-word as unknown (i.e., "false alarm"). The participants were provided with the following instructions in their L1 (see Table 4).

Table 4

Known Word Density Task Instructions

Instructions: (original)	Instructions: (English translation)
下列有幾個句子。 請不必了解句意，但請將下列句子中"所有不會"的生字標記（用滑鼠將生字點選成綠色）。	Below you will see several sentences. Do not read the sentences for comprehension, but simply scan the sentences and click on "all the words that you do not know" (i.e., click on the unknown word to highlight it green).

An analysis of the participants' known word density for the three testing passages was conducted, and the results are reported in Table 5. Descriptive statistics showed that the mean known word density for the three groups ranged from the lowest mean score of 89.79% to the highest mean score of 93.16% among all testing passages. The mean known word density for the pre-test, post-test, and delayed post-test passages was 90.17%, 91.79%, and 91.42%, respectively. Both the Unassisted RR and Assisted RR groups showed an increase of known word density from the pre-test to the post-test, with the Unassisted RR group showing 1.49% increase and the Assisted RR group showing 1.08% increase. The Unassisted RR group increased another 1.5% on the delayed post-test, which summed up to a total of 3% increase from the pre-test to the delayed post-test.

Table 5

Descriptive Statistics for Known Word Density

Group	Test	Mean (%)	SD
Unassisted RR (n=14)	Pre-test	90.17	4.88
	Post-test	91.66	4.00
	Delayed post-test	93.16	4.46
Assisted RR (n=15)	Pre-test	89.96	4.68
	Post-test	91.04	3.22
	Delayed post-test	89.79	5.86
Control (n=13)	Pre-test	90.39	5.09
	Post-test	92.80	4.77
	Delayed post-test	91.43	5.04

Additionally, the participants' known word density of the pre-test passage was used to identify their reading level prior to receiving the treatment, as well as to check whether there was pre-treatment equivalence of the three groups. The pre-test known word density was 90.17% for the Unassisted RR group, 89.96% for the Assisted RR group, and 90.39% for the Control group. A one-way ANOVA revealed that there was no

significant difference (*p=.973*) in the three groups' mean known word density on the pre-test. This suggests that there was baseline equivalence of the groups with regard to the known word density of the pre-test passage.

3.3.3.2 Selection of target words. To generate a pool of target words, 80 content words were initially pre-selected by the researcher – in consultation with their English instructor – from the reading material. These words were chosen according to two criteria: words that the participants would most likely have no prior knowledge of and those that are not in the most frequent 2,000 word-families in the British National Corpus word lists. This was to decrease the chance of the participants having prior knowledge of the words. Additionally, none of the target words appeared in the post-test sessions. The researcher also consulted with the participants' English course instructor on the selection of the reading material, vocabulary difficulty of the passages, and selection of target words. The same text and tasks were piloted on five randomly selected students who did not participate in this study but were from the same grade level and school. They were asked to identify all unknown words in the entire story. The results from identification of unknown words not only indicated the known word density of the text but also helped generate a pool of potential target words.

The pre-selected words generated a pool of potential target words, from which the researcher used an algorithm via the Internet software to randomly select the target words for each of the vocabulary test tasks (see the Instruments section below for a description of the tasks). Twenty items and nine distractors were initially included in each test task.

To finalize the list of target words for the post-tests and data analysis, the current study participants' self-identified unknown words in the reading passages, as well as their

answers on the vocabulary pre-test, were used and analyzed. The words that were known

by many participants were excluded from data analysis. Additionally, words that are

polysemous for which the participants provided an alternative meaning were eliminated

from the analysis.

Table 6

Target Words

Task 1:					
inquire (v.)	x1	rag (n.)	x1	hearth (n.)	x1
urge (v.)	x1	whisk (v.)	x1	slanting (adj.)	x1
knot (n.)	x5	attic (n.)	x7	pantry (n.)	x1
scramble (v.)	x2	paw (n.)	x4	sparrow (n.)	x1
cradle (n.)	x1	yeast (n.)	x1	pattering (adj.)	x1
Distractors: consulted, brewer, contrary, wainscot, rummage, anxious, fetch, exclaimed, snuff					

Task 2:					
oblige (v.)	x1	scoop (v.)	x1	squirm (v.)	x1
distracted (adj.)	x1	mischief (n.)	x2	sneeze (v.)	x1
bundle (n.)	x1	shutter (n.)	x1	stuffy (adj.)	x1
tumble (v.)	x1	jagged (adj.)	x1	trundle (v.)	x1
dreadfully (adv.)	x3	peep (v.)	x1	disarrange (v.)	x1
Distractors: critically, flour, nonsense, crackle, bonnet, determined, lump, plaster, soot					

Task 3:					
distinctly (adv.)	x1	slate (n.)	x2	rasping (adj.)	x1
persuaded (adj.)	x1	apron (n.)	x1	wriggle (v.)	x3
indigestible (adj.)	x2	dough (n.)	x10	gnaw (v.)	x1
poke (v.)	x2	squeak (v.)	x1	unruly (adj.)	x1
chatter (v.)	x1	queer (adj.)	x1	scutter (v.)	x1
Distractors: barrel, candle, fender, yelping, singe, dragging, whip, kettle, cobweb					

Task 4:					
heel (n.)	x1	chimney (n.)	x10	wade (v.)	x1
squeeze (v.)	x1	pastry (n.)	x1	grope (v.)	x1
boldly (adv.)	x1	smear (v.)	x1	saucer (n.)	x1
choke (v.)	x1	attentively (adv.)	x1	mew (v.)	x5
jar (n.)	x3	ledge (n.)	x1	smut (n.)	x1
Distractors: scarcely, wringing, infested, muffin, flue, enormous, consequence, rafter, lath					

The finalized target words (15 items for each vocabulary test task) for data analysis and distractors (nine for each task) are listed in Table 6. Additionally, information on each target word's part of speech is also provided, along with its frequency of appearance. Of the 60 target words, 42 (or 70%) of the words appeared once and 18 (or 30%) of the words appeared at least twice in the text. The target words appeared a total of 29 times for Task 1, 18 times for Task 2, 29 times for Task 3, and 30 times for Task 4.

3.3.3.3 Instruments. The instruments used in this study included reading comprehension tests, vocabulary acquisition tests, learning journals, a background questionnaire, self-evaluation forms, and exit surveys.

3.3.3.3.1 Reading comprehension test. In this study, reading comprehension referred to the ability to understand written texts; it was operationalized as the amount of text information recalled and, accordingly, was measured by immediate written recall protocol. The intermediate-level EFL participants were asked to recall in their L1 immediately after the final reading and were not allowed to go back to the previous web page to refer to the text. They were provided with the following instructions in their L1 (see Table 7). The English translation of the instructions is also provided for the reader's reference.

Table 7

Reading Comprehension Test Instructions

Instructions: (original)	**Instructions: (English translation)**
您對本段文字了解多少？ 請用中文將您所記得的文章內容打下來。**內容越詳盡越好。** 請用『中文』描述。您有 15 分鐘的時間作答。	What did you understand from the text? Please describe what you remember from the text – in as much detail as possible. Please write in Chinese. You have 15 minutes to complete this task.

3.3.3.3.2 Vocabulary acquisition test. In this study, vocabulary acquisition was defined as the knowledge of form-meaning mappings of words that were acquired incidentally and operationalized as the ability to recognize (i.e., receptive knowledge) and/or produce (i.e., productive knowledge) words in context and isolation (Han & Chen, 2010). To measure their incidental vocabulary acquisition, the participants completed the following four tasks (adapted from Han and Chen, 2010). Five minutes were allotted for each of Tasks 1 and 2, and ten minutes were allowed for each of Tasks 3 and 4. The tasks are described in the order they were administered.

(1) Task 1 – Isolated production of words: Task 1 attempted to gauge the participants' ability to produce words in isolation. Here the goal was to determine whether they could link the L2 form of the target word with its L1 meaning (Webb, 2005). The participants were given the L1 meaning of the target word and were asked to provide its L2 form. In order to decrease the chance of the learner providing another L2 form with the same L1 meaning, the first letter of each test item was provided; additionally, its part of speech was provided in parentheses. For example, the participant saw "酵母 *(名詞)=> y*_____" and would need to provide the corresponding L2 word form (i.e., "*yeast*") in the blank. The instructions and example provided to the participants are as follows (see Table 8). The English translation of the instructions is also provided for the reader's reference.

(2) Task 2 – Isolated recognition of words: Task 2 measured the participants' ability to recognize words in isolation. Here the goal was to determine whether they could link the L1 meaning of the target word with its L2 form. The participants were given the L2 word form and were asked to provide its meaning in L1. Additionally, the

word's part of speech was provided in parentheses. For instance, the participant read

"*sneeze (動詞) => _____*" and would need to provide the corresponding L1 meaning

(i.e., "*打噴嚏*") in the blank. They were provided with the following instructions and

example (see Table 9). The English translation of the instructions is also provided for the

reader's reference.

Table 8

Isolated Production of Words Test Instructions

Instructions: (original)	Instructions: (English translation)
請將所提供的中文字義**翻譯成相對應的英文單字**。 *例如：* *快樂 (形容詞)* *h =>* **happy** . ＊註：作答時請留意單字的詞性。 已提供英文單字的第一個字母作為參考。 您有 5 分鐘的時間作答。	Please provide the English word based on the Chinese definition given. *For example:* *快樂 (Adjective)* *h =>* **happy** . * Please note the part of speech provided in parenthesis. The first letter of the word is provided. You have 5 minutes to complete this section.

Table 9

Isolated Recognition of Words Test Instructions

Instructions: (original)	Instructions: (English translation)
請將所提供的英文單字**翻譯成對應的中文意思**。 *例如：* *happy (形容詞)* *=>* *快樂* . ＊註：作答時請留意單字的詞性。 您有 5 分鐘的時間作答。	Please provide the Chinese meaning of the English word given. *For example:* *happy (Adjective)* *=>* *快樂* . * Please note the part of speech provided in parenthesis. You have 5 minutes to complete this section.

(3) Task 3 – Contextual production of words: Task 3 intended to shed light on the participants' productive word knowledge, or their ability to produce words in context. Again, the goal here was to determine whether they could link the L2 form of the target word with its L1 meaning. The paragraphs from which the target words appeared were provided; the target words were left blank and the first letter and L1 meaning of the word were given and highlighted. In addition, the part of speech of each vocabulary test item was provided. The participants were asked to provide the L2 form of the target word. The participants were provided with the following instructions (see Table 10). A sample test item for Task 3 – including the paragraph and question item – is presented below. Again, the English translation of the instructions is also provided for the reader's reference.

Table 10

Contextual Production of Words Test Instructions and Sample Item

Instructions: (original)	Instructions: (English translation)
請在下列段落裡的空格處，將所提供的中文字義**翻譯成相對應的英文單字**。 ＊註：作答時請留意單字的詞性。 英文單字的第一個字母已提供作為參考。 您有 10 分鐘的時間作答。	For the following paragraphs, please fill in the blanks **with appropriate English words** based on the Chinese meaning given. * Please note the part of speech provided in parenthesis. The first letter of the word is provided. You have 10 minutes to complete this task.
Sample test item: "A rolling-pin?" said Ribby. "Did we not hear a roly-poly noise in the attic when we were looking into that chest?" Ribby and Tabitha rushed upstairs again. Sure enough the roly-poly noise was still going on quite **d_____ (清楚地;清晰地)** under the attic floor. "This is serious, Cousin Tabitha," said Ribby. "We must send for John Joiner at once, with a saw." d _____ (副詞) => ｜ **distinctly** ｜.	

(4) Task 4 – Contextual recognition of words: Task 4 measured the participants'
receptive vocabulary knowledge, or their ability to recognize the word, in context. Again,
the goal here was to determine whether the learners could link the L1 meaning of the
target word with its L2 form. The paragraphs from which the target words appeared were
provided, with the target words highlighted, and the participants were asked to provide
the L1 meaning of the target words. Also, the part of speech of each vocabulary test item
was provided. The participants were provided with the following instructions (see Table
11). A sample test item for Task 4 – including the paragraph and question item – is
presented below. Again, the English translation of the instructions is also provided for the
reader's reference.

Table 11

Contextual Recognition of Words Test Instructions and Sample Item

Instructions: (original)	**Instructions: (English translation)**
請寫出下列段落裡所標示的英文單字的**對應中文意思**。 ＊註：作答時請留意單字的詞性。 您有 10 分鐘的時間作答。	For the following paragraphs, please provide **the Chinese definition** of the highlighted English word in each paragraph. * Please note the part of speech provided in parenthesis. You have 10 minutes to complete this section.
Sample test item: Now this is what had been happening to Tom Kitten, and it shows how very unwise it is to go up a **chimney** in a very old house, where a person does not know his way, and where there are enormous rats. Tom Kitten did not want to be shut up in a cupboard. When he saw that his mother was going to bake, he determined to hide. chimney (名詞) =>　　煙囪　　．	

The vocabulary items were ordered differently for the three tests. Additionally, each test was administered about two months apart to account for testing effect.

3.3.3.3.3 Learning journal. Toward the end of each treatment session, the participants were asked to complete a learning journal. They were asked to write in their L1 and to reflect on their learning during the session according to a list of categories provided. The list was adapted from categorical findings from Gorsuch and Taguchi's (2010) qualitative study and included: reading process; reading comprehension and reading speed; the text; the reading instruction; use of reading strategies; motivation to read; language skills development; focus/allocation of attention; and other thoughts. To fit the purpose of the present study, the following categories were added: the participants' perceptions of the words in the reading passages and, for the Assisted RR group, the use of the audio model. The participants were provided with the following instructions (see Table 12).

Table 12

Learning Journal Instructions

Instructions: (original)
『學習日誌』 請回想您剛剛的閱讀和學習，在此學習日誌裡寫出**所想到的任何想法**。 請依據下列所提供的方向進行作答。 請用『中文』作答，內容越詳盡越好。 您有 10-15 分鐘的時間作答。
Instructions: (English translation)
[**Learning Journal**] Please reflect on your reading and learning and write **whatever thoughts come to mind**. Please answer according to the categories provided. **In Chinese**, please write as much as possible. You have 10-15 minutes for this task.

3.3.3.3.4 Background questionnaire. Detailed information about the participants was collected via the online background questionnaire that was administered through Qualtrics (see Appendix B for the original and English translated versions of the questionnaire). Information collected included the following categories: gender, age, first language(s) (L1s), past and current language usage at home and at school, initial exposure to English, past and current English instruction history, recent amount of contact with English (at school and outside of school context), reading behaviors, and perception of the English language and their L1/L2 proficiency. The questionnaire was provided in the participants' L1 to ensure the intermediate-level EFL students understood the questions. Also, it was administered to ensure homogeneity of the participants (i.e., their past English instruction and age) and to see if any modifications to the study needed to be made at the preparation stage.

3.3.3.3.5 Self-evaluation form. For each session, the participants were checked to see whether they had encountered the passage prior to the session. In addition, they were asked to rate their perceived difficulty level of the text – for the first reading and the fifth (final) reading – on a 5-point Likert scale ("1" being "*very easy,*" "2" being "*easy,*" "3" for "*about the right level (neither easy nor hard),*" "4" being "*hard,*" and "5" being "*very hard*"). Also, they were asked to rate their level of understanding of the passage on a scale of 0% to 100%. Furthermore, they were asked to rate their level of on-task-ness on a 5-point Likert scale, from the first to the fifth reading for each passage, when circling unknown words and writing the learning journal during the treatment sessions, and when performing the comprehension and vocabulary tests tasks during the testing sessions.

3.3.3.3.6 Exit survey. After the post-test, the participants were asked to complete an open-ended survey form asking whether they were engaged in any type of self-learning during the study period for any content of the study, their reflections on the reading approach, the reading passage, the recall task, and the vocabulary test task, and lastly, their comments about the study. The survey was provided towards the end of the post-test session, in that memories of the study may have decayed after a two-month summer vacation. The same survey was given again at the end of the delayed post-test.

3.3.3.3.7 Research log. During and immediately following each session, the researcher recorded noticeable incidents and onsite observations that may yield additional insights. Things that were jotted down in the research log included: a few cases of computer troubleshooting or the participants' reported technical issues (e.g., typo, clicking the wrong answer, or mistakenly inserted an answer for the wrong item); the participants' change of perspectives to the text; the participants who were particularly focused during the study; the participants who completed the session requirements substantially earlier or later than others; the participants reporting to the researcher that the text was easy or too hard; the participants' physical and emotional state; any observed violations to the study (e.g., look-ups of the answers on the Internet); date of absences and make-up sessions; the participants' special inquiries; and discussions with their advisor and the research assistant about the participants and the data collection process.

3.3.3.4 Procedure. During the preparation stage, the researcher obtained permission to conduct the study from the school and gathered informed consent from the participants and their parents. The participants were asked to complete an extensive online background questionnaire. Additional information about the participants' English

proficiency and learning were collected, including their six English midterm and final exams scores from the whole academic year.

At the beginning of each session, the participants were first provided with brief instructions in their L1 about the session. Throughout the study, all written and oral descriptions of the procedure were provided in the participants' L1 (i.e., Mandarin Chinese) to ensure that: (1) they understood the procedures and what they needed to complete, given their levels of English proficiency, and (2) none of the students received different amounts of additional target-language input. Also, the participants were informed and constantly reminded that their answers would not affect their English course grades, that they were required to provide honest answers, and that they were not allowed to consult the dictionary or the Internet or to discuss with any other participants about any content of the study during the entire study period. The descriptions were stated as instructions in the Qualtrics online platform.

The method of data collection was Internet-based and presented to the participants through the Qualtrics platform in a computer lab setting. Students' responses and the amount of time they spent on each task were automatically recorded through the platform. The Qualtrics platform also helped pace and sequence the procedure of the study. The data collection system was set up so that students had to answer all questions before they can move on to the next page, to ensure that all questions were answered. Participants were instructed to type an "*X*" for all the questions that they were unable to provide an answer to. They were not allowed to go back to the previous page – to prevent them from changing their answers and revisiting the passage while performing the recall task. In the

very few cases where the participants made a technical mistake, they reported it to the researcher, who then manually corrected their errors from the researcher's end.

Throughout the study period, the participants were not allowed to access the reading materials. Additional efforts were made to ensure that the participants did not review or engage themselves in any type of self-learning of any content of the study. At the beginning of the study, the participants were informed that the reading materials and test results would be made available to them after the study concluded so that they would be able to grasp further learning opportunities.

For the treatment and testing sessions, each participant performed the following procedures:

3.3.3.4.1 Pre-test sessions. In order to reduce priming and fatigue effects, the pre-tests were purposely administered over three days within the same week (Sessions 1-3; Week 1). The vocabulary test Tasks 1 and 2 were administered on Day 1, vocabulary test Tasks 3 and 4 on Day 2, and the reading comprehension test on Day 3.

Step 1 – Vocabulary acquisition test: The participants completed a vocabulary pre-test, with the goal being to gauge their knowledge of the target words prior to the study. Five minutes were allotted for each of Tasks 1 and 2, and ten minutes were allowed for each of Tasks 3 and 4.

Step 2 – Identification of unknown words: Then the participants were asked to identify all unknown words for the reading passages. They were instructed to not read the sentences for comprehension, but to simply scan the sentences and click on all the unknown words. (See the Materials section for a detailed description on identifying unknown words.)

Step 3 – Repeated reading: For each testing session, the participants silently read the passage five times, with a 30-second break in between each reading. They were instructed to read quickly, while trying to understand the passage as much as possible. All participants received the following procedural instructions (see Table 13).

Table 13

Repeated Reading Task Instructions

Instructions: (original)	Instructions: (English translation)
請閱讀下列的文章。	Please read the following passage.
此段落您將會閱讀『5次』。 請快速閱讀，但同時也**盡可能的理解文意**。	You will read the passage "5 times". Please read quickly, and at the same time, try to **understand the passage as much as possible**.

Step 4 – Reading comprehension test: Following the final re-reading, immediate written recall of the text was conducted. The participants had 15 minutes to complete the task.

Step 5 – Self-evaluation: Lastly, the participants completed the self-evaluation form to report on whether they had read the passage before, as well as rate the level of text difficulty and their level of on-task-ness for each re-reading and each task performance.

3.3.3.4.2 Treatment sessions. During each treatment session (Sessions 4-11; Weeks 2-5), the Unassisted RR and Assisted RR groups participants completed the following steps. The Control group, on the other hand, did not engage in any of the treatment sessions.

Step 1 – Identification of unknown words: (The participants received the same procedural instructions as those in the pre-test sessions.)

Step 2 – Repeated reading: The Assisted RR and the Unassisted RR groups silently read the same passage five times, with a 30-second break between each reading. (See the Pilot Studies section for a detailed description on the rationale behind the number of re-readings.) The participants were instructed to read quickly, and at the same time, to try to understand the passage as much as possible. The Unassisted RR group received the same procedural instructions as those in the testing sessions, as described previously in *Step 3 – Repeated reading* of the pre-test sessions. The Assisted RR group participants re-read the passage while listening to the audio model recording that was uniformly operated and broadcasted by the researcher. Before starting to play the audio recording, the researcher signaled to the participants that the next audio play was about to begin. The Assisted RR group participants were provided with the following procedural instructions (see Table 14).

Table 14

Assisted Repeated Reading Task Instructions (Treatment Sessions)

Instructions: (original)	Instructions: (English translation)
請跟著**播放的聲音檔，同步閱讀**下列的英文文章。	Please **read the following passage while listening** to the audio recording broadcasted.
此段落您將會 聽＋閱讀『5次』。 請快速閱讀，但同時也**盡可能的理解文意**。	You will read and listen to the passage "5 times". Please read quickly, and at the same time, try to **understand the passage as much as possible**.

Step 3 – Learning journal: At the end of each session, the participants completed the learning journals in their L1 to reflect on their learning during the session.

Step 4 – Self-evaluation: (Participants received the same procedural instructions as those in the pre-test sessions.)

3.3.3.4.3 Post-test sessions. Upon finishing the treatment sessions, the participants proceeded to the testing sessions (Sessions 12-14; Week 6), which took place over three days within the same week. On the first day, the participants completed the first two vocabulary test tasks, the second day the third and fourth vocabulary test tasks, and on the third day the reading comprehension test. The procedure of the post-test was the same as that of the pre-test, which included: (1) a vocabulary test, (2) identification of unknown words, (3) repeated reading, (4) reading comprehension test, and (5) self-evaluation. The vocabulary test tasks and items were the same as those in the pre-test, and the target words in each task were the same as those in the pre-test but ordered differently. After the post-test, the participants completed the exit survey.

3.3.3.4.4 Delayed post-test sessions. Two months after the post-test, the participants returned to the computer lab for the delayed post-test sessions (Sessions 15-17; Week 16). In between the immediate and delayed post-tests, the participants went on summer vacation and were away from regular academic curriculum. As the participants were entering the final year of vocational high school, they were fully occupied with preparing for license tests over the two-months summer vacation and were busy at the school factory practicing their machinery and automatic control skills. The procedure of the delayed post-tests was the same as that of the previous testing sessions. Again, the vocabulary test tasks and items were the same as those in the pre-test and the post-test, but the items in each task were ordered differently. Before concluding the study, the participants completed the exit survey to reflect on and provide comments about the entire study.

3.3.4 Data Coding

The data from the reading comprehension and the vocabulary acquisition tests from the pre-test, post-test, and delayed post-test sessions were coded by three raters. Additionally, the three testing passages were divided into idea units, and each idea unit was categorized into hierarchical levels and assigned different scores by nine raters. Inter-rater agreement was computed by comparing the raters' coding and scoring of the data and then calculating the percentage of items that were agreed upon by all raters.

3.3.4.1 Reading comprehension tests. Prior to the data collection, the three testing passages were first divided into idea units. Upon completion of the data collection, the participants' written recall protocols from the three testing sessions were coded for the amount of correct idea units recalled from the text. The researcher conducted pre-coding norming sessions and post-coding discussion sessions with the raters to code the testing passages and the participants' recall protocols.[2] When discrepancies in the coding occurred, the raters discussed until an agreement was reached.

3.3.4.1.1 Testing passages coding. Prior to the data collection, the pre-test, post-test, and delayed post-test passages were first divided into idea units by nine English native speakers who were also doctoral students and/or instructors in ESL. Adapting Han and Peverly's (2007) definition, an idea unit constitutes a phrase, clause, or sentence – in lieu of a single word – and provides a new unit of information that "matches part of the

---

[2] Due to schedule conflict, the norming sessions and the post-coding discussion sessions were carried out in different small groups, in lieu of having all coders meet at the same time. The researcher attempted to serve as a facilitator during the sessions. However, researcher's subjectivity most likely had inadvertently influenced the results. Also, the researcher may have had an influence on decision-making when discrepancy occurred. The role that the researcher played in the coding and decision making process is a limitation of the study and should be noted.

overall message of the reading passage" (p. 28); for example, "*there was an old cat.*"
Adapting Carrell (1985), each separate idea unit consisted of: a single clause (main, subordinate, adverbial, relative, noun); infinitival construction; gerundive; nominalized verb phrase; conjunct (that connects phrases and clauses); adverbial phrase; participle phrase; or prepositional phrase. Adverbial phrase (e.g., "*all over the house*"), participle phrase (e.g., "*hidden in the chimney*"), and noun clause (e.g., "*that he had not time*") were added to Carrell's original coding scheme, as they pertain to the passages for the present study. In a few cases where the sentences are fairly short (for example, "*She looked in the pantry*" or "*They went into the barn*") and ideas become too fragmented if further divided into even smaller units, the prepositional phrase was not coded as a separate idea unit, and the sentence was categorized as a main clause and remained one single unit. In a few instances, adding a word in front of a phrase could turn one grammatical construction into another; when such discrepancies occurred between the raters, the coders' judgment came into play and they discussed until an agreement was reached. Choices and judgment for the units were made based on consistency of level of information for all idea units, and based on which is more suitable for the overall message. The pre-test, post-test, and delayed post-test passages resulted in a total of 49, 56, and 50 idea units, respectively. The scoring rubric is presented in Appendix D with coding of idea units for the testing passages.

In addition to dividing the testing passages into idea units, the present study categorized the idea units into two hierarchical levels: Main-idea (M) and Detail (D). "Main-idea (M)" idea units represent major ideas or main topics of the passage (Carrell, 1985). They express a basic and fundamental idea of the text and typically form the

summary of the text and provide information that is essential and indispensable to the text (Osa-Melero, 2015). "Detail (D)" idea units represent minor ideas or subtopics of the passage (Carrell, 1985). They provide support, evidence and/or justification to the main points, and the information is somewhat necessary for rebuilding a summary of a text (Osa-Melero, 2015).

The testing passages were further coded into Main-idea (M) and Detail (D) idea units by the same nine raters. The pre-test passage resulted in a total of 21 Main-idea and 28 Detail idea units; the post-test passage resulted in a total of 20 Main-idea and 36 Detail idea units; and the delayed post-test passage resulted in a total of 24 Main-idea and 26 Detail idea units. The total number of idea units for each testing text and the number of different level idea units are presented below (see Table 15).

Table 15

Number of Idea Units

	Pre-test	Post-test	Delayed post-test
Main idea (M)	21	20	24
Details (D)	28	36	26
Total number of idea units	49	56	50

In order for the scoring scheme to be more precise and construct valid, the present study assigned different scores to the idea units to reflect their "weight" and relative importance to the whole story and context (Berkemeyer, 1989; Well, 1986). A full score of 2 was given for accurately recalling "Main-idea" idea units and a score of 1 for "Detail" idea units. Paraphrasing or alternative wording was considered correct as long as the same meaning was conveyed. To account for partial credits, a half-point score was given,

which resulted in a score of 1 for major ideas and 0.5 for details that were recalled partially correct. A score of zero was assigned to incorrect responses or a missing/no response (see Table 16).

Table 16

Idea Units Scoring

	Totally correct	Partially correct	Incorrect or Missing
Main idea (M)	2	1	0
Details (D)	1	0.5	0

Thus, the maximum possible raw score for each of the testing texts was 70 (=21*2+28*1), 76 (=20*2+36*1), and 74 (=24*2+26*1), respectively; the minimum raw score for each passage was 0 (see Table 17).

Table 17

Idea Units Maximum and Minimum Raw Scores

	Pre-test	Post-test	Delayed post-test
Maximum raw score	70	76	74
Minimum raw score	0	0	0

3.3.4.1.2 Recall protocols coding. Upon completion of the data collection, the participants' written recall protocols from the three testing sessions were coded (by three raters, i.e., the researcher and two English native speakers who were also doctoral students and instructors in ESL) for the amount of correct idea units recalled from the text. Specifically, the protocols were divided into idea units and then scored for accuracy of each idea unit from the passage (Carrell, 1985). There were instances of inference-type responses (for example, providing the answer "Tom went all over the house" for the

storyline *Tom went up the staircases to the attic and down to the basement*); when this occurred, the participants' inference-type responses were categorized into either Main-idea or Detail idea units, based on the relative importance of the response to the passage. Each unit was scored chronologically, and decisions were made based on the answer's level of completeness, in which case a total or partial score was given (Bernhardt, 1992). Because the maximum raw scores varied slightly for the three passages, the raw score for each passage was transformed into a percentage of the total idea units by dividing the raw score by the maximum possible score in the original passage (Carrell, 1985). The inter-rater agreement of the protocols was 90.8%.

3.3.4.2 Vocabulary acquisition tests. Tasks 1 and 2 were coded for productive and receptive knowledge of the target words in isolation, and tasks 3 and 4 were coded for productive and receptive knowledge of the target words in context. As vocabulary learning is incremental in nature and a learner's word knowledge (as well as each individual type of word knowledge) ranges on a continuum – from zero to partial to precise – rather than being known versus unknown dichotomy (Schmitt, 2010), the scoring scheme should reflect the incremental nature of the development, or depth, of vocabulary knowledge.

To reflect the participants' depth of word knowledge on a continuum, two points were awarded for a correct response, 1 point for a partially correct response, and 0 points for an incorrect response. Providing answers in either the original (stem) form or derivational form of the target words was coded as correct responses; for instance, two points were assigned for providing either "*smut*" or "*smuts*" for the noun "*smuts*" in the passage, and for providing either "*poke*" or "*poked*" for the verb "*poked*." Partial credit

was given for providing an approximate meaning for the target word, indicating that the participants have some knowledge of the word. For example, for the target word "*peep,*" the participant provided "偷偷的 *[furtive]*" in lieu of "偷看 *[looking furtively at something]*." Spelling errors were coded as partially correct responses; for instance, if the participant wrote "*inquare*" for the target word "*inquire,*" one point was given. Even though spelling was not the main focus of this task, a response containing spelling errors still suggests having a partial knowledge of the word form and was thus considered partially correct. As the finalized target word list consisted of 15 items per test task, the maximum score was 30 points for each task; the total score for all four tasks combined was 120 points. Each participant's raw score was transformed into a percentage score – by dividing the raw score by the total 120 points – for further analyses.

The data was coded independently by three raters, and the inter-rater agreement was 97.8%.

3.3.5 Data Analysis

Several rounds of data analysis were conducted. An initial round of analysis was administered to eliminate participants that did not meet the pre-set criteria. The participants were eliminated from data analysis if they belonged to any of the following criteria. Participants who reported to have been under intensive contact with English outside of school during the study period, or who had lived in an English-speaking environment for an extended period of time, were eliminated. Participants were checked for familiarity (i.e., have content familiarity) with each reading passage in the self-evaluation form provided at the end of each session. Those that reported to have read the story prior to the study, and thus have prior knowledge and content familiarity, were

excluded. If a subject missed a study session, he or she was asked to make up the session within a day or two. Participants who did not complete all required components of the study, due to change of school or hospitalization (i.e., attrition), were also excluded. Five participants were identified to have special needs (i.e., learning disability, autism, or school phobia) and were eliminated. The participants were also not considered for data analysis if they did not follow the instructions (i.e., violation of the rules): for example, if they were found or self-reported to have looked up the meanings of the words or L1 translation of the story, peeked at other participants' answers during a testing session, or resorted to a website for any information related to the study. This was to ensure that the effects are truly the result of the treatment and not contaminated by external factors or resources. Those that identified statistically significantly more unknown words from the pre-test passage than the majority of the participants were excluded from data analysis. Additionally, the participants who scored statistically significantly different on the pre-test (i.e., outliers) based on SPSS results were removed from the remainder of the analysis to ensure that the three groups are comparable. The number of participants that were included for the remainder of the data analysis for each group was: Unassisted RR (n=14), Assisted RR (n=15), and Control (n=13) group.

Next, to conduct a quantitative analysis, statistical analyses were performed using SPSS V.24 for each group's performance on both the reading comprehension and vocabulary acquisition assessment measures. The two independent variables were the type of learning conditions (Unassisted RR, Assisted RR, and Control groups) and time of the assessment (pre-test, post-test, and delayed post-test), and the continuous dependent variables were the participants' performance on the comprehension tests and

vocabulary tests. Thus, a two-way mixed ANOVA was run to determine if the scores changed differently over time depending on which treatment condition each group underwent.

The following describes the procedure for conducting statistical analyses. In particular, the data were tested for outliers and normality, homogeneity of variances, homogeneity of covariances, and sphericity, and analyses were conducted for group and time using two-way mixed ANOVA method to determine if there was a two-way interaction between the two independent variables (i.e., with group as the between-subjects factor and time as the within-subjects factor) on the continuous dependent variable (i.e., test scores). If any assumption was not met for running two-way mixed ANOVA, extra steps were taken to ensure that the violation was dealt with (see below for details on how these were done), and the results should be interpreted with caution.

As a first step, the data were tested for outliers using boxplots generated when running the two-way mixed ANOVA procedure. This was to eliminate any negative impact that outliers may cause on the two-way mixed ANOVA by "distorting the differences between cells of the design" and "causing problems when generalizing the results (of the sample) to the population" (Laerd Statistics, 2016). The boxplots for the pre-test scores were inspected for values greater than 1.5 box-lengths from the edge of the box, which was an indication of an outlier.

Another step was taken to ensure baseline equivalence of the groups prior to the treatment. Groups are considered equivalent if their differences (i.e., group mean differences) are less than or equal to one-quarter of the pooled standard deviation, regardless of statistical significance (Ho et al., 2007; National Center for Education

Evaluation and Regional Assistance, 2014). Therefore, the study controls for the participants' pre-test scores on which the groups differ to ensure pre-intervention differences were not greater than one-quarter of the pooled standard deviation. Participants whose pre-test scores did not comply with the statistical standard were eliminated from the analysis to establish baseline equivalence so as to provide an adequate comparison.

If an outlier was found, the participant was removed from the remainder of the analysis in order to ensure pre-treatment equivalence of the groups and to prevent the individuals from having such an undue influence on the generalization of the results. After deleting the outlier, the data were re-run to generate the SPSS Statistics output for Outliers, Normality, Variance and Covariance, Sphericity, and the main two-way mixed ANOVA result. After ensuring baseline comparability of the groups, the assumption of homogeneity of variance, assumption of sphericity, and assumption of equality of covariances, were tested and reported as part of the two-way mixed ANOVA procedure.

Levene's test of homogeneity of variance was used to test the assumption of homogeneity of variances; i.e., whether the variances of the dependent variable are equal between the groups of the between-subjects factor (Laerd Statistics, 2016). If the variances were unequal, which can affect the Type I error rate, a more conservative simple error was used to deal with this problem.

Homogeneity of covariances (multisample sphericity) was tested using Box's test of equality of covariance matrices. If the assumption of covariances was not met, the violation of the assumption was not so much an issue, as there was a similar number of

participants in each group and the following analysis was separated into separate repeated

measures ANOVAs for each group (Laerd Statistics, 2016).

Mauchly's test of sphericity was used to test whether the assumption of sphericity

was met (i.e., whether the variance of the differences between any combination of levels

of the within-subjects factors was equal). Whether a statistically significant two-way

interaction exists between group and time was determined by looking at descriptive

statistics and estimates, profile plots, and two-way mixed ANOVA test results. If the

assumption of sphericity was violated (i.e., $p < .05$) for the two-way interaction, a

correction test of Greenhouse-Geisser estimates was used to determine the result of the

two-way group*time interaction. Specifically, results generated from the two-way

ANOVA procedure were checked to see if there was a statistically significant two-way

interaction between group and time on the test scores; that is, whether the effect of the

treatment for the groups was different and the test scores changed differently for each

group across time.

If the two-way interaction was not statistically significant, results from the post-

hoc test of the main two-way mixed ANOVA procedure with a Greenhouse-Geisser

correction were examined to determine (1) if the data has any statistically significant

main effects for the between-subjects factor (i.e., main effect of group; that is, testing for

differences in test scores between the different groups "collapsed" across time, or simply

comparing the three groups regardless of the time point) or (2) if the data has any

statistically significant main effects for the within-subjects factor (i.e., main effect of

time; that is, testing for differences in test scores "collapsed" across group, or simply

comparing the three time points regardless of the intervention group). If a main effect

was statistically significant for either of the factors (i.e., groups or time), post hoc tests were run to follow up the main effect with pairwise comparisons, and the results will inform us where the differences in test scores between time points and/or groups lie.

If there was a statistically significant two-way interaction, simple main effects were run for group and time (Laerd Statistics, 2016). To test the simple main effects for "time," three separate within-subjects ANOVAs (i.e., repeated measures ANOVAs) were performed to test for differences in test scores between time points for each of the between-subjects factor, i.e., group. In testing for the simple main effects for "group," three separate between-subjects ANOVAs (i.e., three separate one-way ANOVAs) were performed to conduct separate tests for differences in test scores between groups at each time point. If a simple main effect was statistically significant for either of the factors, post hoc tests were run to follow up the simple main effect with pairwise comparisons, and the results will inform us where the differences in test scores lie.

In addition to the statistical analyses, the participants' data were also analyzed from a qualitative standpoint. Learners' comments from each of the testing and treatment sessions were considered for the purpose of triangulation, as well as to gain a micro-level perspective on any effects of the treatment and other factors on the participants' performance. Specifically, data collected from the participants' learning journals were analyzed to see if any patterns emerged or if the data yielded additional insights. Also, data collected from the background questionnaire, self-evaluation forms, and exit surveys were analyzed. First, the participants' recorded responses were read multiple times to catch any emerging patterns and outstanding topics or themes. After categorizing the comments into topics (which included reading comprehension, vocabulary acquisition,

the use of the audio model, and re-reading of the passage), the participants' responses

were further coded according to a detailed subcategory (i.e., "Response Category"). Data

garnered primarily from the learning journal, as well as surveys and questionnaires, may

help elucidate participants' subjective perspective during the study and supplement the

statistical analysis with a more in-depth learner perception. The results of the data

analyses are presented in the next chapter.

Chapter IV

RESULTS

This chapter describes the results of the study. First, descriptive statistics for the

reading comprehension and vocabulary acquisition tests scores are provided. Additionally,

the results of the effects of the treatment over time and group on the participants'

comprehension and vocabulary acquisition scores are presented. The latter sections of

this chapter aim to explore the data with the goal to gain a deeper understanding of the

relationship between the independent variables and changes in learner behavior. In doing

so, the participants' perceptions garnered from the learning journals and surveys and any

patterns of intra-learner variation were placed under critical examination.

As mentioned in Chapter III, an initial round of analysis was administered to

eliminate the participants that did not meet the pre-set criteria. Analysis of the data

showed that 17 participants from the Unassisted RR group, 16 from the Assisted RR

group, and 15 from the Control group met all the criteria and were included for a

preliminary set of data analysis.

4.1 Reading Comprehension

The first research question asks whether L2 repeated reading promotes reading

comprehension. Following the final re-reading of the passages for the pre-test, post-test,

and delayed post-test sessions, the study measured the participants' reading

comprehension via immediate written recall protocol. As noted in Chapter III, given that

the maximum possible raw scores varied slightly from text to text (i.e., 70, 76, and 74,

respectively, for the pre-test, post-test, and delayed post-test), the participants' raw scores were transformed into a percentage of the total number of idea units to make the test scores comparable. To conduct a quantitative analysis, a two-way mixed ANOVA was performed using SPSS V.24 for each group's comprehension pre-test, post-test, and delayed post-test scores.

A preliminary set of data analyses was first conducted on the participants' comprehension pre-test scores to check for outliers and baseline equivalence of the groups. The results showed that for the comprehension pre-test, the Unassisted RR group had the highest mean score of 14.41% (SD=16.87), followed by the Control group with a mean score of 11.90% (SD=14.07) and then the Assisted RR group with a mean score of 6.65% (SD=8.65). The boxplots and histogram of the preliminary set of data analysis on the participants' comprehension pre-test scores are illustrated in Figure 1 and Figure 2, respectively. An inspection of the boxplots for the pre-test scores indicated outliers.

In addition, the data were checked for baseline equivalence of the groups on the comprehension pre-test to ensure that pre-intervention differences were less than or equal to one-quarter of the pooled standard deviation. The results showed that the group mean differences (i.e., Unassisted RR – Assisted RR = 7.76; Unassisted RR – Control = 2.51; Control – Assisted RR = 5.25) between the Assisted RR and the Unassisted RR or the Control groups were larger than one-quarter of the pooled standard deviation (i.e., 3.43), which suggests that the pre-test scores of the three groups were not equivalent and baseline equivalence of the groups was not met.
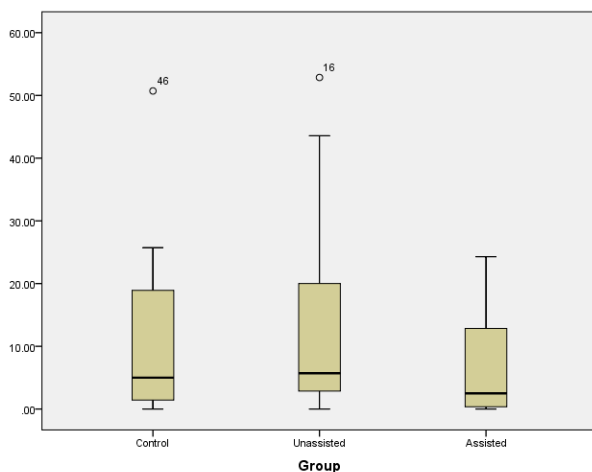
Figure 1

Boxplots for Comprehension Pre-test Scores (Preliminary)
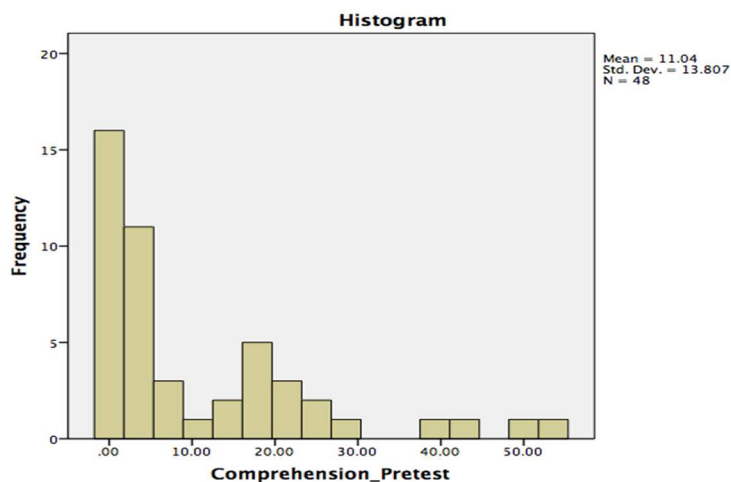
Figure 2

Histogram for Comprehension Pre-test Scores (Preliminary)

Participants whose pre-test scores did not comply with the statistical standards (for both the comprehension and vocabulary tests scores) were eliminated from the remainder of the study to establish baseline equivalence so as to provide an adequate comparison. Based on the boxplots (Figure 1), the participants that were eliminated had a comprehension score value above 30(%) on the pre-test, indicating a higher proficiency at the starting point (i.e., outlier). Another step was taken to ensure baseline equivalence

of the groups by removing participants until all group mean differences were less than or equal to one-quarter of the pooled standard deviation (for both the comprehension and vocabulary tests scores). As a result, six participants (two from the Control group, three from the Unassisted RR group, and one from the Assisted RR group) were eliminated in order to ensure comparability of the groups and to prevent the individuals from having such an undue influence on the generalization of the results. The participants that met all the assumptions for both the comprehension and vocabulary scores remained and their data were analyzed and discussed. The following numbers of participants from each group were included for the remainder of the data analysis and for discussion: Unassisted RR (n=14), Assisted RR (n=15), and Control (n=13) group.

After eliminating the outliers and ensuring baseline equivalence of the groups, the data were re-analyzed by SPSS to generate the Statistics output. The results showed that on the comprehension pre-test, the Unassisted RR group had the highest mean score of 7.91% (SD=9.14), followed by the Control group with a mean score of 7.86% (SD=8.47) and then the Assisted RR group with a mean score of 6.14% (SD=8.70). All the indices now better comply with the assumptions of the parametric statistical tests. The standard deviations for all three groups decreased and showed less variation than those of the preliminary results. There were no outliers in the data for the pre-test scores; additionally, the histogram now better fits the assumptions of normality (Figure 3). Baseline equivalence of the groups was met for the comprehension pre-test scores, as the group mean differences (Unassisted RR – Assisted RR = 1.77; Unassisted RR – Control = .05; Control – Assisted RR = 1.71) were less than one-quarter of the pooled standard deviation (i.e., 2.19). Lastly, the Test of Between-subjects Effects indicated that there was

no statistically significant difference for the three groups on the comprehension pre-test (*p=.934*).

Histogram

Mean = 7.26
Std. Dev. = 8.603
N = 42
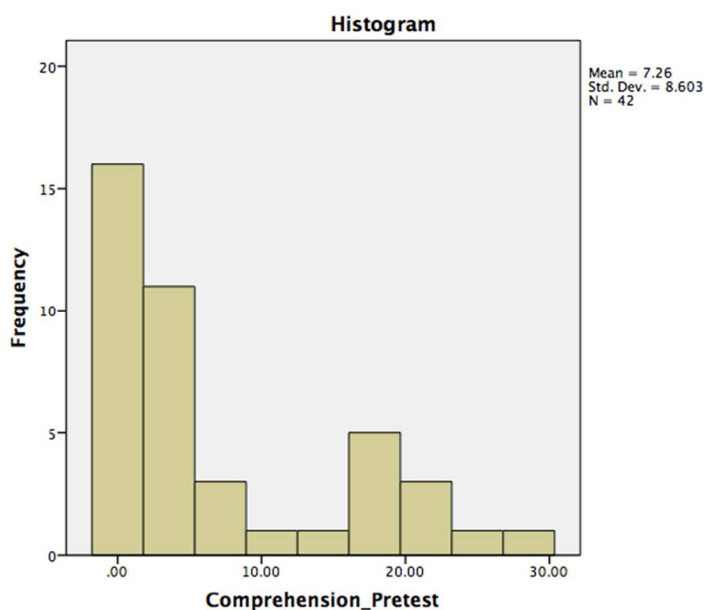
Figure 3

Histogram for Comprehension Pre-test Scores

Next, the assumption of homogeneity of variance, the assumption of equality of covariances, and the assumption of sphericity, were tested as part of the two-way mixed ANOVA procedure. The results are presented below.

The results showed that the assumption of homogeneity of variances was met (i.e., *p>.05*) for the comprehension pre-test (*p=.989*), post-test (*p=.501*), and delayed post-test (*p=.929*) as assessed by Levene's test of homogeneity of variance, suggesting that the variances of the dependent variable (i.e., comprehension test scores) are equal.

The assumption of homogeneity of covariances (i.e., multisample sphericity) was tested using Box's test of equality of covariance matrices. The results showed that the data have met (i.e., *p>.05)* the assumption of homogeneity of covariances, as the test was not statistically significant (*p=.600*).

Mauchly's test of sphericity was used to test whether the assumption of sphericity was met (i.e., whether the variance of the differences between any combination of levels of the within-subjects factors was equal). The results from the Mauchly's test of sphericity indicated that the assumption of sphericity was violated (i.e., p<.05) for the two-way interaction, $\chi^2(2)=6.865$, $p = .032$, which suggests that the variance of the difference was not equal. Thus, a correction test of Greenhouse-Geisser was used to estimate this adjustment and to determine the result of the two-way group*time interaction.

Table 18

Descriptive Statistics for Reading Comprehension Tests

Group	Test	Mean (%)	Mean Gain (%)	Standard Deviation
Unassisted RR (n=14)	Pre-test	7.91[a]	–	9.14
	Post-test	4.09[b]	-3.82	4.36
	Delayed post-test	4.30[c]	0.21	6.05
Assisted RR (n=15)	Pre-test	6.14	–	8.70
	Post-test	4.74	-1.40	6.66
	Delayed post-test	3.56	-1.18	5.22
Control (n=13)	Pre-test	7.86	–	8.47
	Post-test	4.00	-3.86	4.15
	Delayed post-test	3.74	-0.26	4.67

a: 7.91(%) = 5.537 (mean raw score) / 70 (maximum score for Pre-test) x 100%
b: 4.09(%) = 3.1084 (mean raw score) / 76 (maximum score for Post-test) x 100%
c: 4.30(%) = 3.182 (mean raw score) / 74 (maximum score for Delayed post-test) x 100%

Table 18 presents the descriptive statistics of the scores obtained on the reading comprehension measure, and Figure 4 illustrates percentage of idea units recalled for the comprehension pre-test, post-test, and delayed post-test. The means for the Unassisted RR group ranged from 4.09% to 7.91%, and the standard deviations ranged from 4.36 to 9.14. The means for the Assisted RR group ranged from 3.56% to 6.14%, and the

standard deviations ranged from 5.22 to 8.70. The means for the Control group ranged

from 3.74% to 7.86%, and the standard deviations ranged from 4.15 to 8.47.
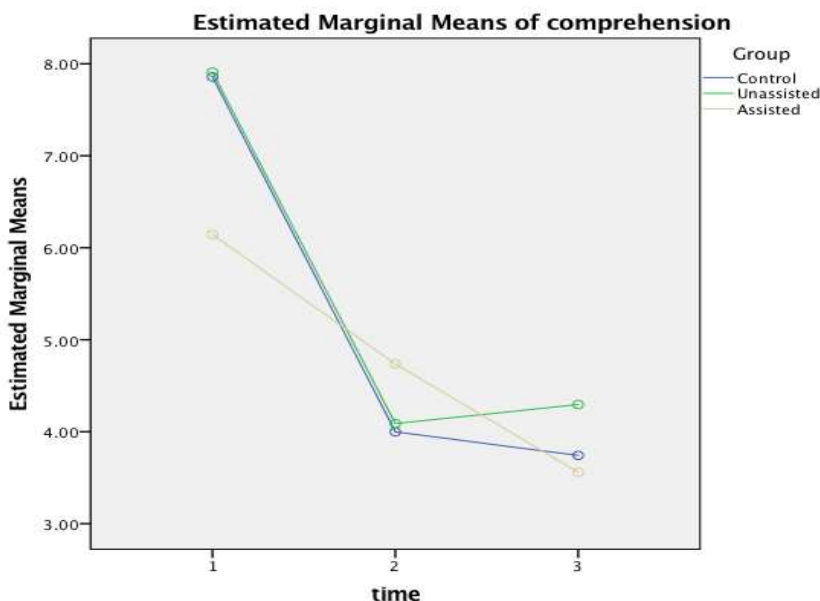
Note: The green line represents the Unassisted RR group, the beige line represents the Assisted RR group, and the blue line represents the Control group. The vertical axe denotes the mean scores (percentage) on the comprehension tests, and the horizontal axe denotes "time" of the assessment: pre-test (Time 1), post-test (Time 2), and delayed post-test (Time 3).

Figure 4

Reading Comprehension Test Scores

On the comprehension pre-test, the Unassisted RR group had the highest mean

score of 7.91% (SD=9.14), followed closely by the Control group with a mean score of

7.86% (SD=8.47) and by the Assisted RR group with a mean score of 6.14% (SD=8.70).

On the post-test, the Control group (-3.86%) and the Unassisted RR group (-3.82%)

showed a larger decrease in the percentage mean score, whereas the Assisted RR group (-

1.40%) displayed some decrease compared to the pre-test. On the delayed post-test, the

Unassisted RR group exhibited a minimal amount of increase (0.21%), and the Assisted

RR group (-1.18%) and the Control group (-0.26%) showed a slight decrease in the mean scores compared to the post-test.

Results indicated that there was a no statistically significant two-way interaction between group and time on the comprehension test scores, $F(3.433, 66.937)=.216$, $p=.907$, partial $\eta^2=.011$, $\varepsilon=.858$. Therefore, results from the post-hoc test of the main two-way mixed ANOVA procedure with a Greenhouse-Geisser correction were examined to determine if the data has any statistically significant main effects for the between-subjects factor (i.e., main effect of group) or the within-subjects factor (i.e., main effect of time).

The main effect of time (that is, testing for differences in test scores "collapsed" across group, or simply comparing the three time points regardless of the intervention group) showed a statistically significant difference in mean comprehension test scores at the different time points, $F(2, 3.433)=4.310$, $p=.017$, partial $\eta^2=.100$. As a main effect was statistically significant for time, post hoc tests were run to follow up the main effect with pairwise comparisons. Results showed that the comprehension test score statistically significantly decreased at post-test compared to pre-test ($M=3.028$, $SE=1.102$, $p=.027$). The score did not statistically significantly decrease from the post-test to the delayed post-test ($M=.409$, $SE=1.184$, $p=1.000$) or from the pre-test to the delayed post-test ($M=3.437$, $SE=1.514$, $p=.086$). Results of the main effect of group (that is, testing for differences in test scores between the different groups "collapsed" across time, or simply comparing the three groups regardless of the time point) showed that there was no statistically significant difference in mean comprehension test scores between the three groups, $F(2, 39)=.066$, $p=.937$, partial $\eta^2=.003$.

In sum, the results indicated that the treatment did not seem to promote reading comprehension scores and that a transfer of practice effect to a new passage in terms of reading comprehension was not found.

4.2 Vocabulary Acquisition

The second research question investigates whether L2 repeated reading leads to incidental vocabulary acquisition. The participants' vocabulary acquisition was measured via the four vocabulary test tasks, each measuring isolated production, isolated recognition, contextual production, and contextual recognition of words. The finalized target word list for data analysis consisted of 15 items per test task with a full score of 2 for each correctly answered item. The maximum score was 30 points for each task, and the total score for all four tasks combined was 120 points. Each participant's raw score was transformed into a percentage score – by dividing the raw score by the total 120 points – for further analyses. To conduct a quantitative analysis, a two-way mixed ANOVA was performed using SPSS V.24 for each group's performance on the vocabulary pre-test, post-test, and delayed post-test.

A preliminary set of data analysis was first conducted on the participants' vocabulary pre-test scores to check for outliers and baseline equivalence of the groups. The results showed that for the vocabulary pre-test, the Assisted RR group had the highest mean score of 5.42% (SD=8.22), followed by the Control group with a mean score of 4.17% (SD=3.62) and then the Assisted RR group with a mean score of 3.77% (SD=3.02). The boxplots and histogram of the preliminary set of data analysis on the participants' vocabulary pre-test scores are illustrated in Figure 5 and Figure 6,

respectively. An inspection of the boxplots for the participants' pre-test scores indicated outliers.
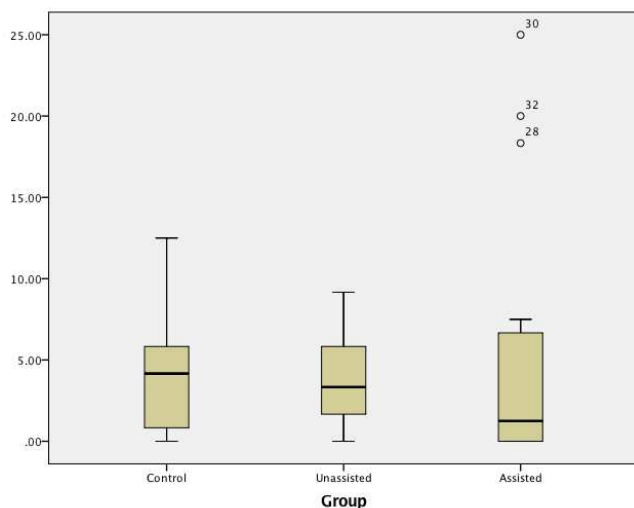
Figure 5

Boxplots for Vocabulary Pre-test Scores (Preliminary)
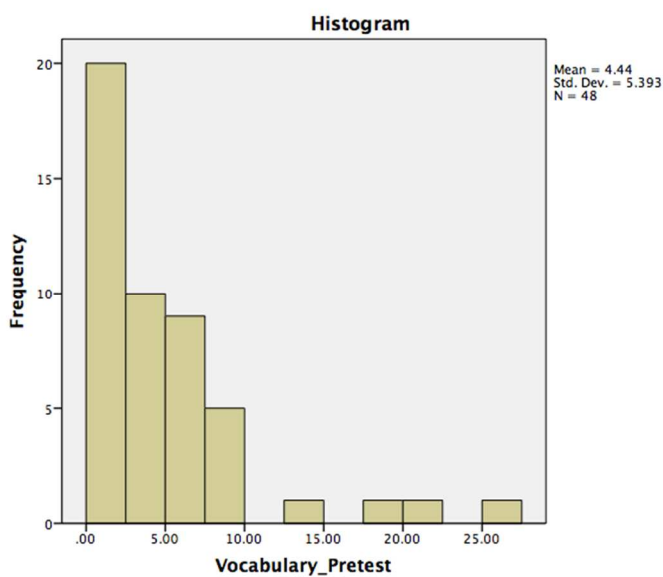
Figure 6

Histogram for Vocabulary Pre-test Scores (Preliminary)

In addition, the data were checked for baseline equivalence of the groups on the vocabulary pre-test to ensure that pre-intervention differences were less than or equal to

one-quarter of the pooled standard deviation. The results showed that the group mean

difference (i.e., Assisted RR – Unassisted RR = 1.64; Assisted RR – Control = 1.25;

Control – Unassisted RR = .39) between the Assisted RR and the Unassisted RR groups

was larger than one-quarter of the pooled standard deviation (i.e., 1.37), which suggests

that the pre-test scores of the groups were not equivalent and baseline equivalence of the

groups was not met.

Participants whose pre-test scores did not comply with the statistical standards

(for both the comprehension and vocabulary tests scores) were eliminated from the

remainder of the study to establish baseline equivalence so as to provide an adequate

comparison. Based on an inspection of the boxplot (Figure 5) for values greater than 1.5

box-lengths from the edge of the box, the participants that were eliminated had a

vocabulary score value higher than a 20(%) on the pre-test, indicating a higher

proficiency at the starting point (i.e., outlier). Another step was taken to ensure baseline

equivalence of the groups by removing participants until all group mean differences were

less than or equal to one-quarter of the pooled standard deviation (for both the

comprehension and vocabulary tests scores). As a result, six participants (two from the

Control group, three from the Unassisted RR group, and one from the Assisted RR group)

were eliminated in order to ensure comparability of the groups and to prevent the

individuals from having such an undue influence on the generalization of the results. The

participants that met all the assumptions for both the comprehension and vocabulary

scores remained and their data were analyzed and discussed. The following numbers of

participants from each group were included for the remainder of the data analysis and for

discussion: Unassisted RR (n=14), Assisted RR (n=15), and Control (n=13) group.

Figure 7

Histogram for Vocabulary Pre-test Scores

After eliminating the outliers and ensuring baseline equivalence of the groups, the data were re-analyzed by SPSS to generate the Statistics output. The results showed that on the vocabulary pre-test, the Assisted RR group had the highest mean score of 4.11% (SD=6.57), followed by the Unassisted RR group with a mean score of 3.87 % (SD=3.23) and then the Control group with a mean score of 3.46% (SD=2.98). All the indices now better comply with statistical assumptions of the parametric statistical tests. The standard deviations for all three groups decreased and showed less variation than those of the preliminary results. There were no outliers in the data for the pre-test scores; additionally, the histogram now better fits the assumptions of normality (Figure 7). A closer inspection of the vocabulary test results indicated that the group mean differences between the Unassisted RR and the Control groups (.41), between the Control and the Assisted RR groups (.65), and between the Unassisted RR and the Assisted RR groups (.24) were all less than one-quarter of the pooled standard deviation (i.e., 1.17), indicating baseline equivalence of the groups was met. Lastly, the Test of Between-subjects Effects indicated

that there was no statistically significant difference for the three groups on the vocabulary pre-test (*p=.828*).

Next, the assumption of homogeneity of variance, the assumption of equality of covariances, and the assumption of sphericity, were tested as part of the two-way mixed ANOVA procedure. The results are presented below.

The results showed that the assumption of homogeneity of variances was met (i.e., *p>.05*) for vocabulary pre-test (*p=.088*) and post-test (*p=.073*), as assessed by Levene's test of homogeneity of variance, suggesting that the data have equal variance for the pre-test and the post-test. However, the assumption of homogeneity of variances was not met (i.e., *p<.05*) for the delayed post-test (*p=.017*), suggesting that the variances of the test scores are not equal on the delayed post-test. If the variances are unequal, this can affect the Type I error rate (Laerd Statistics, 2016); in the present study, this was overcome by using a more conservative simple error.

The assumption of homogeneity of covariances (i.e., multisample sphericity) was tested using Box's test of equality of covariance matrices. The results showed that the data did not have homogeneity of covariances, as the test was statistically significant (*p<.001*). Even though the assumption of covariances was not met, the violation of the assumption is not so much of an issue, as there were similar numbers of participants in each group (i.e., by having equal groups). As there was no homogeneity of covariances, the following analysis was separated into separate repeated measures ANOVAs for each group (Laerd Statistics, 2016).

Mauchly's test of sphericity was used to test whether the assumption of sphericity was met (i.e., whether the variance of the differences between any combination

of levels of the within-subjects factors was equal). The results from the Mauchly's test of

sphericity indicated that the assumption of sphericity was not met (i.e., *p<.05*) for the

two-way interaction, $\chi^2(2) = 25.594, p = .000$; this suggests that the variance of the

difference was not equal. Thus, a correction was made to correct for this and the

Greenhouse-Geisser was used to estimate this adjustment. A correction test of

Greenhouse-Geisser estimates was used to determine the result of the two-way

group*time interaction, i.e., whether the two-way interaction is statistically significant.

Table 19

Descriptive Statistics for Vocabulary Acquisition Tests

Group	Test	Mean (%)	Mean Gain (%)	Standard Deviation
Unassisted RR (n=14)	Pre-test	3.87[a]	–	3.23
	Post-test	12.20[b]	8.33	9.73
	Delayed post-test	7.74[c]	-4.46	5.95
Assisted RR (n=15)	Pre-test	4.11	–	6.57
	Post-test	6.61	2.50	7.26
	Delayed post-test	5.06	-1.55	6.62
Control (n=13)	Pre-test	3.46	–	2.98
	Post-test	4.74	1.28	5.09
	Delayed post-test	3.24	-1.50	2.15

a: 3.87(%) = 4.644 (mean raw score) / 120 (maximum score for the Pre-test) x 100%
b: 12.20(%) = 14.640 (mean raw score) / 120 (maximum score for the Post-test) x 100%
c: 7.74(%) = 9.288 (mean raw score) / 120 (maximum score for the Delayed post-test) x 100%
(Note. maximum possible raw score =15 items x 4 test tasks x 2 points full score for a correct response = 120; minimum possible raw score = 15 items x 4 test tasks x 0 points for an incorrect response = 0)

Table 19 presents the descriptive statistics for the scores obtained on the

vocabulary acquisition measure, and Figure 8 illustrates the percentage of vocabulary test

scores for each group on the pre-test, the post-test, and the delayed post-test. The means

for the Unassisted RR group ranged from 3.87% to 12.20%, and the standard deviations

ranged from 3.23 to 9.73. The means for the Assisted RR group ranged from 4.11% to

6.61%, and the standard deviations ranged from 6.57 to 7.26. The means for the Control

group ranged from 3.24% to 4.74%, and the standard deviations ranged from 2.15 to 5.09.
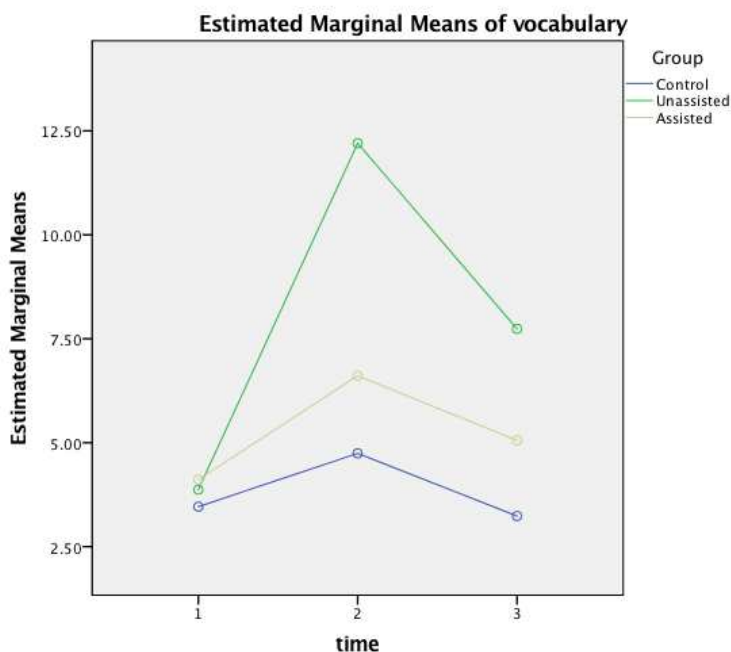
Note. Figure 8 depicts the percentage of correctness for each group for the vocabulary
pre-test, post-test, and delayed post-test. The green line represents the Unassisted RR
group, the beige line represents the Assisted RR group, and the blue line represents the
Control group. The vertical axe denotes the mean scores (percentage) on the vocabulary
tests, and the horizontal axe denotes "time" of the assessment: pre-test (Time 1), post-test
(Time 2), and delayed post-test (Time 3).

Figure 8

Vocabulary Acquisition Test Scores

On the vocabulary pre-test, the Assisted RR group had the highest mean score of

4.11% (SD=6.57), followed closely by the Unassisted RR group with a mean score of

3.87% (SD=3.23) and by the Control group with a mean score of 3.46% (SD=2.98). The

Unassisted RR group showed a larger mean gain (8.33%) from the pre-test to the post-

test, whereas the Assisted RR group displayed some gains (2.50%) and the Control group

exhibited a minimal amount of increase (1.28%) on the post-test. On the delayed post-test,

the Unassisted RR group exhibited some amount of decrease (4.46%), and the Assisted

RR group (1.55%) and the Control group (1.50%) showed a slight amount of decrease

from the post-test.

Results indicated that there was a statistically significant two-way interaction

between group and time on the vocabulary test scores, $F(2.684, 52.346)=3.771$, $p=.019$,

partial $\eta^2=.162$, $\varepsilon=.671$. As there was a statistically significant interaction, simple main

effects were run for group and time to test for differences in test scores between groups at

each category of the within-subjects factor, time, and to test for differences in test scores

between time points for each category of the between-subjects factor, group (Laerd

Statistics, 2016).

To test the simple main effects for "time," three separate within-subjects

ANOVAs (i.e., repeated measures ANOVAs) were performed to test for differences in

vocabulary test scores between time points for each of the between-subjects factor, i.e.,

group. Results of the simple main effect for time suggest that there was a statistically

significant effect of time on vocabulary test scores for the Unassisted RR group, $F(2,$

$26)=8.429$, $p=.002$, partial $\eta^2=.393$. For the Unassisted RR group, the vocabulary test

score statistically significantly increased at the post-test compared to the pre-test

($M=8.33$, $SE=2.376$, $p=.012$), suggesting that the treatment elicited statistically

incidental vocabulary gains from the pre-test to the post-test. Likewise, the vocabulary

test score was statistically significantly higher at the delayed post-test compared to the

pre-test $(M=3.869, SE=.963, p=.004)$, and the score was not statistically significantly

different between the post-test and the delayed post-test ($M=4.464, SE=2.409, p=.260$),

suggesting that the statistically significant amount of vocabulary gains on the post-test was retained.

Similarly, there was a statistically significant effect of time on vocabulary test scores for the Assisted RR group, $F(2, 28)=7.131$, $p=.003$, partial $\eta^2=.337$. For the Assisted RR group, the vocabulary acquisition score statistically significantly increased at the post-test compared to the pre-test ($M=2.500$, $SE=.801$, $p=.023$), indicating significant gains in incidental word knowledge after the treatment. The score statistically significantly decreased from the post-test to the delayed post-test ($M=1.556$, $SE=.563$, $p=.046$), and there was no statistically significant difference between the delayed post-test and the pre-test ($M=.944$, $SE=.619$, $p=.447$), suggesting that the significant improvement on the post-test due to the treatment did not sustain.

On the other hand, the results did not show a statistically significant effect of time on vocabulary test scores for the Control group, $F(2, 24)=1.146$, $p=.325$, partial $\eta^2=.087$. As shown in Table 19, the participants of the Control group had a minimal increase (1.28%) of mean score from the pre-test to the post-test and a minimal decrease (1.50%) from the post-test to the delayed post-test.

In testing for the simple main effects for "group," three separate between-subjects ANOVAs (i.e., three separate one-way ANOVAs) were performed to conduct separate tests for differences in vocabulary test scores between groups at each time (i.e., test) point. Results of the simple main effects for group suggest that there was no statistically significant difference in vocabulary test scores between groups on the pre-test, $F(2, 39)=.068$, $p=.934$, partial $\eta^2=.003$.

For the vocabulary post-test, there was a statistically significant difference in the test scores, *F(2, 39)=3.538, p=.039, partial η²=.154.* The vocabulary test score was statistically significantly higher (*M=7.459, SE=2.94, p=.040*) for the Unassisted RR group compared to the Control group. The Assisted RR group's test score was not statistically significantly higher (*M=1.868, SE=2.90, p=.796*) than that of the Control group, and it was not statistically significantly lower (*M=5.5913, SE=2.84, p=.134*) than that of the Unassisted RR group.

Lastly, results of the simple main effect for group suggest that there was no statistically significant difference in vocabulary test scores between groups on the vocabulary delayed post-test, *F(2, 39)=2.403, p=.104, partial η²=.110.*

In sum, the results suggested that the treatment led to statistically significant incidental vocabulary gains from the pre-test to the post-test for both the Unassisted RR and the Assisted RR groups, but retention of the significant gains was found only for the Unassisted RR group on the delayed post-test. Additionally, the Unassisted RR group showed a greater increase in scores from the pre-test to the post-test than the Assisted RR group, and a statistically significant difference was found only between the Unassisted RR group and the Control group on the post-test.

4.3 Supplementary Findings

In addition to examining changes to the participants' test performance prior to and following their undergoing of the treatment, further analysis was conducted on the participants' responses on the learning journals, survey, and questionnaires. Data garnered primarily from the learning journal, as well as surveys and questionnaires, may

help elucidate participants' subjective perspective during the study and supplement the aforementioned statistical analysis with a greater understanding of the learners' thoughts and experiences than the statistics alone would provide. Data from the same participants that were included for the remainder of the quantitative analysis were analyzed (i.e., 14 participants from the Unassisted RR group, 15 from the Assisted RR group, and 13 from the Control group). The results of the analyses are presented below. Specifically, the participants' perceptions are presented in the following categories: reading comprehension, vocabulary acquisition, the use of the audio model, and re-reading of the passage.

Table 20 and Table 21 summarize the participants' perception of their reading comprehension for the Unassisted RR group and the Assisted RR group, respectively. The comments ranged from indicating total lack of comprehension to partial comprehension or understanding of the general plot of the story. Several participants (16 total counts from the Unassisted RR group, and 17 total counts from the Assisted RR group) commented that they were able to understand the main ideas of the text. On the other hand, more participants (36 counts for the Unassisted RR group, and 39 counts for the Assisted RR group) reported that they could not or do not quite understand the text. This corroborated with the quantitative finding that the text was generally difficult for the participants and their comprehension test scores were at the lower end.

Also, a few participants (three total counts from the Unassisted RR group, and five total counts from the Assisted RR group) commented on the text being a little difficult. Although not the focus of the present study, a few participants (five counts from the Unassisted RR group, and one count from the Assisted RR group) also commented on

the positive effect of re-reading on their reading fluency; the participants expressed that they were able to notice an increase in their reading speed, even though they did not notice any improvement in their understanding of the text. On the other hand, a few participants (six counts from the Unassisted RR group, and one count from the Assisted RR group) were able to sense their own improvement in the understanding of the text.

Table 20

Perception of Reading Comprehension (Unassisted RR Group)

Response category	Session								Total count
	1	2	3	4	5	6	7	8	
I can understand the general plot.	3	2	3	2	1	1	3	1	16
I'm starting to sense improvement in my understanding of the text.	1	2			2			1	6
There's no improvement in my understanding of the text, but I can read faster.	1	1		1				2	5
I don't quite understand the text.	3	5	5	3	3	2	1	1	23
I don't understand the text.		1	4	2	1		2	3	13
The text is a little difficult.		1			1	1			3

Table 21

Perception of Reading Comprehension (Assisted RR Group)

Response category	Session								Total count
	1	2	3	4	5	6	7	8	
I can understand the general plot.	5	3	2	2	2	1	1	1	17
I'm starting to sense improvement in my understanding of the text.		1							1
There's no improvement in my understanding of the text, but I can read faster.	1								1
I don't quite understand the text.	2	2	2	1			1	2	10
I don't understand the text.	3	3	4	7	4	4	2	2	29
The text is a little difficult.	1	1	1				1	1	5

Table 22 and Table 23 illustrate the participants' perception of the words and vocabulary acquisition for the Unassisted RR group and the Assisted RR group, respectively. As garnered from the tables, many participants in the Unassisted RR group commented that they gradually understood the unknown words through repeated reading (nine counts) and that the words were becoming easier (nine counts). This corroborated with the quantitative results as reflected by an increase of the vocabulary scores for the Unassisted RR group. A participant in the Unassisted RR group commented that he was able to understand most of the text by the second reading, but he could not understand some portions of the text because of the unknown words. Unlike the Unassisted RR group, the Assisted RR group participants did not mention that they gradually understood the unknown words, and only one comment expressed that the words became easier through repeated reading. This supported the quantitative results in that the Unassisted RR group showed higher vocabulary gains and retention than the Assisted RR group.

However, for both groups, more participants expressed that they thought the words were difficult (17 counts from the Unassisted RR group, and 8 counts from the Assisted RR group) and that they could not understand many of the words (19 counts from the Unassisted RR group, and 28 counts from the Assisted RR group) or some of the words (18 counts from the Unassisted RR group, and nine counts from the Assisted RR group). The participants' general impression of word difficulty corroborated, again, with their higher portion of unknown words in the texts.

Also, three participants in the Unassisted RR group specifically noted that some unknown words remained unknown, unless the reader was provided with contextual cues. Some (seven counts) of the participants from the Assisted RR group requested in their

learning journals for the meaning or definition of difficult words to be provided along

with the reading materials.

Table 22

Perception of Vocabulary Acquisition (Unassisted RR Group)

Response category	Session								Total count
	1	2	3	4	5	6	7	8	
Through repeated reading, I gradually understand the unknown words.		2	1	1	2	1	1	1	9
The words are (becoming) easier.	2	1	2	1	1	1		1	9
The words are difficult.	2	2	1	3	3	2	1	3	17
I don't understand many of the words.	2	5	3	3	1	1	3	1	19
I don't understand half of the words.									0
I don't understand some of the words	3	1	1	2	2	4	2	3	18
I use the words before and after to guess the meaning of an unknown word.	3	1	2	2	1	1	1		11
Unknown words remain unknown.	1	2							3

Table 23

Perception of Vocabulary Acquisition (Assisted RR Group)

Response category	Session								Total count
	1	2	3	4	5	6	7	8	
The words are (becoming) easier.	1								1
The words are difficult.			2	1		1	2	2	8
I don't understand many of the words.	5			5	5	3	6	4	28
I don't understand half of the words.	2						1		3
I don't understand some of the words	3			1	1	1	1	2	9
I use the words before and after to guess the meaning of an unknown word.	1								1
It would be great to provide word meaning.	2	2	1	2					7

Lastly, the participants commented that they used the words before and after the

unknown word to guess its meaning. Interestingly, more participants from the Unassisted

RR group (11 counts) reported doing so than the Assisted RR group (one count). A

reason for this may be that the Unassisted RR group participants were able to read at their

own pace and were thus able to use this metacognitive strategy when encountering unknown words, whereas the Assisted RR group participants were reading along with the audio model and did not have time to do so.

In order to obtain retrospective information about the participants' perception of the usage of the audio model while reading, the Assisted RR group participants were asked to comment on their thoughts of the audio model. As shown in Table 24, overall the comments revealed a positive attitude toward the usage of the audio model in terms of its effectiveness on their learning. Many participants thought that listening to the audio model deepened their impression (original words: "*加深印象*") of the content of the story (eight counts) and was helpful and useful (16 counts); also, they commented that the usage of the audio model was "very good" (27 counts) and the reading of the passage was very clear (eight counts). The participants' comments generally yielded a positive attitude toward the use of the audio model during repeated reading, which corroborates with the literature.

Interestingly, one participant commented on Sessions 1 through 5 that the speed of the audio reading was too fast for him, but this impression went away from Session 6 on. This may have resulted from his lack of comprehension of the passages, as the comment corroborated with his scores on the comprehension tests (a score of 0% on all three tests). Although his vocabulary test scores showed a minimal increase from a score of 2% on the pre-test to 11% for both the post-test and delayed post-test, he commented on the evaluation survey following the post-test that there were too many unknown words which prevented him from understanding the text.

Table 24

Perception of Audio Model Usage (Assisted RR Group)

Response category	Session								Total count
	1	2	3	4	5	6	7	8	
Deepened my impression of the content.	1	1	1	1	1	1	1	1	8
Helpful and useful.	2	2	2	2	2	2	2	2	16
Very good.	5	4	3	3	3	3	3	3	27
Suitable for my level.	0	0	0	0	0	1	0	0	1
Good, but I don't understand the content.	1	0	0	0	0	0	0	0	1
Not helpful because I couldn't understand the content.	1	0	1	0	0	0	0	0	2
Reading of the passage (words and prosody) was very clear.	1	1	1	1	1	1	1	1	8
Prosody was too dramatic for me.	1	0	0	0	0	1	0	0	2
The reading speed was too fast.	1	1	1	1	1	0	0	0	5

The participants were asked to reflect on their learning in relation to reading the

passage multiple times. Table 25 and Table 26 summarize the participants' responses for

the Unassisted RR group and Assisted RR group, respectively.

For the Unassisted RR group, the participants generally considered repeated

reading as beneficial for their understanding of the passage. Many participants noted that

re-reading of the passage helped them understand the passage better (22 counts), helped

with their understanding of the main idea (three counts), and deepened their impression

of the content (one count). They also rated re-reading as "good or not bad" (six counts)

and believed re-reading was meaningful, beneficial, and helpful (four counts). Some

participants noticed that re-reading helped increase their reading speed (11 counts). One

participant started to understand the unknown words through re-reading. However, some

commented that they did not understand the content and still could not understand after

multiple readings (three counts); others noted that they still did not understand some of the words (five counts). Some participants thought the process was tedious (eight counts) and that there were too many reading times (seven counts), and that re-reading was not helpful but meaningless (four counts).

Table 25

Perception of Repeated Reading (Unassisted RR Group)

Response category	Session								Total count
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
RR deepened my impression of the content.	1								1
RR helped me understand the passage better.	2	3	2	3	4	4	2	2	22
RR helped my understanding of the main idea.	1	2							3
Good; not bad.			2	2	1	1			6
Re-reading is meaningful, beneficial and helpful.			2			1		1	4
RR helped increase my reading speed.	1	2	2	3		1		2	11
I started to understand the unknown words through re-reading.	1								1
I didn't understand the content and still couldn't understand after re-reading.			1		1	1			3
I still didn't understand some of the words.			1			1	2	1	5
The process was tedious.	2	1	1		1	3			8
There were too many reading times.	2	1	1	1	1	1			7
Not helpful; meaningless.	1	2			1				4

A participant commented that he started to understand the passage after the third or fourth reading. Another participant even reported that through re-reading, he started to understand the unknown words and later realized that he misunderstood part of the story. Additionally, a participant wrote that starting from the third re-reading, he began to pick up the details, i.e., information that he did not realize during earlier readings. Interestingly, a participant reported that re-reading helped him understand better, but he

felt that additional support was still needed to really understand the story [*original words:* "能加強理解，不過我覺得不懂的地方還需要別人指示才能真正清楚*"*].

As shown in Table 26, the Assisted RR group participants generally had positive thoughts of re-reading the passages. They commented that re-reading deepened their impression (eight counts) and memory (three counts) of the content, as well as helped their listening skills (eight counts), understanding of the passage (nine counts), and reading speed (two counts). Two participants were able to sense their own improvement. Also, they rated the use of re-reading as "good, or not bad" (19 counts) and noted that re-reading was meaningful, beneficial, and helpful (eight counts). However, what they found hard to understand remained unknown after multiple readings (five counts).

Table 26

Perception of Repeated Reading (Assisted RR Group)

Response category	Session								Total count
	1	2	3	4	5	6	7	8	
RR deepened my impression of the content.	1	1	1	1	1	1	1	1	8
RR helped my memory (of the content).	1	0	1	0	0	0	0	1	3
RR helped my listening skill and deepened my aural memory traces.	1	1	1	1	1	1	1	1	8
I could sense my own improvement.	1						1		2
RR helped me understand the passage better.	1	1	1	1	1	1	0	1	7
RR helped my understanding of the main idea.	0	0	1	0	0	1	0	0	2
Good; not bad.	2	5	0	3	3	2	2	2	19
Re-reading was meaningful, beneficial and helpful.	1	1	1	1	1	1	1	1	8
RR helped increase my reading speed.	0	0	1	0	0	0	1	0	2
I didn't understand the content and still couldn't understand after re-reading.	1	1	1	1	0	1	0	0	5
The process was tedious.	0	0	1	0	0	0	0	0	1
There were too many reading times.	0	0	0	0	0	1	0	0	1

Furthermore, the participants commented in the learning journal that the process was tedious (one count) and that there were too many reading times (one count); a participant reported in Session 6 that reading the text three times was sufficient. Interestingly, fewer participants from the Assisted RR group (i.e., one count each) provided such comments than the Unassisted RR group (i.e., seven and eight counts).

Even though one of the participants did not perform well on the reading comprehension and vocabulary acquisition test scores, he commented throughout the treatment sessions that the repeated exposure to a text enhanced his reading comprehension. The other participant reported throughout the entire treatment phase that repeated exposure to the same text, along with simultaneous aural input, enhanced his listening skills and deepened traces of aural memories. In three journal entries, he expressed that through repeated exposure, our ears will remember the sounds we hear [*original words: 耳朵會記起聲音 (Session 3), 這是耳熟能詳的道理 (Session 4), 耳朵會有記憶 (Session 5)*]".

4.4 Summary of Main Findings

To conduct statistical analyses, a two-way mixed ANOVA was run to determine the effect of the treatment over time and group on the participants' comprehension and vocabulary acquisition scores, respectively.

The results indicated that the treatment did not seem to promote reading comprehension; there was a statistically significant decrease in scores from the pre-test to the post-test for all groups, suggesting that a transfer of practice effect to a new passage was not found. The treatment led to statistically significant incidental vocabulary gains

from the pre-test to the post-test for both the Unassisted RR and the Assisted RR groups, but retention of the vocabulary gains was found only for the Unassisted RR group on the delayed post-test. Additionally, the Unassisted RR group showed a greater increase in scores from the pre-test to the post-test than the Assisted RR group, and a statistically significant difference was found only between the Unassisted RR group and the Control group on the post-test.

Lastly, the analysis of the data gathered through the learning journals, surveys, and questionnaire revealed that the participants' comments generally corroborated and supported the quantitative results. Specifically, many participants from both the Unassisted RR and the Assisted RR groups commented that they could not or did not quite understand the text, which corroborated with their low comprehension scores, but some of the participants were able to sense their own improvement in understanding of the text through re-reading. Some participants in the Unassisted RR group reported that they gradually understood the unknown words and showed usage of metacognitive strategies, which supported the Unassisted RR group's higher gains in vocabulary scores from the pre-test to the post-test. For both groups, the participants generally thought that the text was difficult, that there were too many unknown words which impeded their understanding of the text, and that some unknown words remained unknown after multiple readings. The Assisted RR group participants generally had a positive attitude toward the usage of the audio model and thought that listening to the audio model deepened their impression of the content. For both groups, the majority of participants deemed re-reading generally beneficial in that it helped them understand better and assisted them in grasping the main ideas; however, some participants expressed that what

they found hard to understand remained unknown after multiple readings and that the

process was tedious and there were too many reading times. The participants' reflections

generally yielded positive perceptions of the benefits of RR with respect to their reading

comprehension, vocabulary acquisition, usage of the audio model, and re-reading of the

passage.

Chapter V

DISCUSSION AND CONCLUSION

This chapter presents a discussion of the main findings. First, the results of the study are discussed in relation to the research questions posed in Chapter II. Additional findings are also discussed, followed by a discussion of the limitations of the study and directions for future research.

5.1 Repeated Reading and Reading Comprehension

The first research question asked whether L2 repeated reading promotes reading comprehension. For five encounters with each text, the participants read consecutive passages of a story with 90% known words. The results indicated that there was a statistically significant decrease (2.97%) in the mean comprehension scores from the pre-test to the post-test and that there was no significant difference in scores between the pre-test and the delayed post-test or between the post-test and the delayed post-test for all three groups combined (see Table 18 and Figure 4 in Chapter IV). In terms of which repeated reading instructional condition (Assisted RR or Unassisted RR) was more beneficial to L2 reading comprehension, the data did not reveal a significant advantage for either of the two treatment groups, and neither of the treatment groups showed a statistically significant difference in scores when compared to the Control group. This was corroborated by the results from the qualitative data for both the Unassisted RR and Assisted RR groups. While both groups yielded some positive attitudes toward repeated reading in benefiting their understanding of the main ideas of the story and providing

them with a sense of improvement in understanding the text, there were twice as many comments throughout the study expressing that they could not or did not quite understand the passages. In sum, the findings suggest that L2 repeated reading did not promote reading comprehension and that a transfer of practice effect to new text was not found.

Although the literature generally suggests that RR is beneficial for reading comprehension, the present study does not support this claim. A major factor that contributes to this finding may be associated with the participants' unknown word density of the reading materials (i.e., around 10% unknown word density, or around 90% known word density). Research has suggested that upwards of 98% of the vocabulary in a text needs to be known for the majority of L2 readers to achieve adequate comprehension of fiction texts, and that a 95% known word density can provide "acceptable" comprehension with some support (e.g., instruction or resources) (Klingner, 2004). In the present study, the participants recognized around 90% of the words, which indicates that the texts were above their current reading level. This was reflected in the qualitative data (e.g., learning journals, self-evaluation): As noted in the previous chapter, many participants expressed that the text was generally difficult for them. Thus, in the case where the participants knew around 90% of the words, the five repeated encounters with the text may not be adequate to promote their reading comprehension.

Additionally, too many unknown words in a text automatically impedes word decoding and leads to reduced fluency. If fluent reading does not take place and reading is slow and laborious, comprehension suffers as a result (Klingner, 2004). This was supported by the qualitative analysis of the participants' responses. Many participants commented that they were not able to understand some or many of the words;

additionally, a few participants noticed that due to many unknown words in the text, it was hard to understand the text and they had to read slowly. As Pressley (2006) explains, "when a reader slowly analyzes a word into component sounds and blends them, a great deal of capacity is consumed, with relatively little left over for comprehension of the word, let alone understanding the overall meaning of the sentence containing the word and the paragraph containing the sentence" (p. 68). If a reader expends too much attention on lower-level processing, less attention is available for higher-level comprehension processing (LaBerge & Samuels, 1974), and consequently, comprehension suffers. Furthermore, if there are too many unknown words in the text, the reader may not be able to guess the meaning from limited information available in the context, and thus, RR alone may not lead to better comprehension (Taguchi et al., 2012).

The participants' slower reading rate may also support why repeated reading did not promote reading comprehension for the present study. Research has suggested a minimum reading rate of 200 wpm for readers to read with full comprehension (Taguchi & Gorsuch, 2002). In the current study, the participants' initial reading rate on the pre-test (ranging from an average of 78 wpm to 121 wpm), the post-test (104 wpm to 173 wpm), and the delayed post-test (84 wpm to 166 wpm) passages were all considerably lower than the minimum rate for readers to read with full comprehension. A slow reader – often attending to every word and reading in isolated units (rather than meaningful sentences) – will likely have an overloaded short-term memory. As a result of not being able to retain enough information in the short-term memory for an extensive period of time, the readers' meaning construction processes of the connected text would most likely have been ineffective (Gorsuch & Taguchi, 2010).

The current study's finding was consistent with Taguchi and Gorsuch (2002), which also used a difficult text and no additional support. Their RR group showed an initial reading rate of 153 wpm on the post-test, and no transfer effects of re-reading practice to a new, unpracticed passage were found. They argued that the readers were not able to free themselves from word recognition tasks, to process the text efficiently, and to use appropriate higher order comprehension skills that would facilitate better performance in comprehension.

Although the literature generally suggests that RR is beneficial for L2 reading comprehension, these claims were actually found in studies that used level appropriate texts or graded readers (i.e., books or passages with reduced range of vocabulary and simplified grammatical structures). These types of text were selected for the purpose of fluency training, and thus may have contributed to better comprehension in these studies. Unlike previous studies, the present study used a more difficult text (i.e., 90% known word density), which may not be conducive to reading comprehension. This could suggest that easier passages may be more beneficial for reading comprehension through RR than difficult texts. This may also suggest that a certain threshold of proficiency (e.g., known word density) is necessary for RR to have its intended effects on promoting reading comprehension.

Additionally, the majority of previous RR-L2 studies provided the participants with external assistance to RR. Research has suggested that reading or RR supplemented with various types of form-focused instruction better promotes text comprehension. Such beneficial effect was also found in Han and Chen (2010), which provided their heritage speaker participant with a variety of form-focused instruction to RR. The results showed

that the participant was able to process passages far beyond her reading level: that is, difficult texts at an average of 75% known word density. While the benefit of RR was found in conjunction with graded readers, level-appropriate texts, and difficult texts provided with the assistance of additional instruction, repeated reading may not work in promoting comprehension when the unknown word density reached 10% and the L2 participants were not provided with additional support. This suggests that RR alone may still be insufficient for EFL learners dealing with difficult texts and that additional support to repeated reading may still be necessary.

Interestingly, a participant reported in the learning journal that through re-reading, he understood the text better, realized his misunderstandings of the text, and began to pick up details of the story, but he believed that additional support was still needed to really understand the story. This finding corroborates the literature regarding the need for additional support when using difficult texts. As mentioned earlier, Hu and Nation (2000) suggested that with 95% known word density, acceptable reading comprehension might still be viable if the reader is provided with some support (e.g., instruction or resources). When the learners are provided with texts beyond their current reading level (as is the case for the present study), such additional support may become even more essential.

There were other divergences between the present study and previous studies that could explain present findings. The majority of previous studies included more treatment sessions (i.e., 20 to 30 sessions). However, the present study had only eight treatment sessions, due to practical constraints. The number of sessions may have been insufficient for creating a conducive learning condition for texts that were above the participants' current level of proficiency. However, more research is needed to determine whether

there would have been a measurable effect for repeated reading, if the participants were provided with more treatment sessions.

In sum, the findings suggest that a certain threshold of proficiency (e.g., known word density) may need to be met before the comprehension benefits of repeated reading could be reaped. Thus, future studies can consider investigating the text difficulty level (threshold), the type and amount of external assistance, and the number of sessions needed, in order to lead to optimal comprehension outcomes.

Another finding of the present study is that there was a statistically significant decrease in reading comprehension scores from the pre-test to the post-test for all three groups combined. One reason for the decrease in scores may be associated with the comparability between the testing passages; namely, the post-test passage may have been slightly more difficult. Despite other factors being controlled for (e.g., number of words, type-token ratio, word difficulty (word level), and mean log frequency) for the testing passages, a post hoc analysis of the texts suggested that the post-test passage was slightly more difficult in terms of the Lexile Measure (the Lexile measure was 920L, 1070L, 800L, respectively, on the pre-test, the post-test, and the delayed post-test passages). This suggests differences in reading demand of the texts in terms of their semantic difficulty and syntactic complexity. Also, for the story of the post-test passage, the setting was constantly changing. This may have increased the participants' cognitive load and increased the difficulty for processing and understanding the text. As a participant in the Assisted RR group commented on the post-test passage, "劇情太亂，場景一直變，看不太懂 [*The plot was in disorder; the setting was constantly changing; [I] don't quite understand [the text]*.]". This is a limitation of the study and should be noted. Another

possible factor for the statistically significant decrease in comprehension scores from the pre-test to the post-test may be attributed to the participants' condition (i.e., level of energy) during the post-test sessions. The post-test sessions took place toward the end of a semester with the final exam period approaching, and the participants may have expended all their energy preparing for the final exams; anecdotally, the researcher observed that the participants came to the post-test sessions looking tired. The participants' energy level may have played an influential role in their performance and should be noted. Despite the limitations, the participants' comprehension scores were similar for the three tests and between the treatment and control groups, suggesting that the treatment did not promote reading comprehension and it did not lead to a transfer of practice effect to new passages.

Research has generally suggested that dual modality reading/RR is more beneficial than single modality reading/RR. However, the present study did not find any statistically significant difference between the dual modality (Assisted RR) and the single modality (Unassisted RR) groups, although the dual modality group showed less of a decline from the pre-test to the post-test. There is one L1 study and one L2 study in the literature that are relevant to the present study's findings. Rogowsky, Calhoun, and Tallal's (2016) study investigated the effects of L1 reading, listening, and dual modality input on reading comprehension, and the participants performed similarly across the three conditions for both the post-test and the delayed post-test. For RR studies in the L2 setting, Liu and Todd (2014) is the only study that compared dual modality repeated reading and single modality repeated reading – and also included a comprehension measure. Even though their results indicated that the dual modality group significantly

outperformed the single modality group, the difference in scores is weakened by their study's limitations (see Chapter II for a discussion of the study and its limitations). Thus, results of the comparison between dual and single modality input remain inconclusive. This suggests the need for more research on the modality of input so that we can come closer to a conclusion in terms of which modality input (single or dual) is more beneficial for L2 reading comprehension through repeated reading.

5.2 Repeated Reading and Vocabulary Acquisition

The second research question investigated whether L2 repeated reading leads to incidental vocabulary acquisition (i.e., form-meaning mappings of words). The reader will be reminded that this measure reflected a composite of scores on four different vocabulary tasks, each measuring isolated production, isolated recognition, contextual production, and contextual recognition of words. Statistical analysis indicated that for the Unassisted RR group, the treatment elicited statistically significant incidental vocabulary gains (i.e., 8.33% increase) from the pre-test to the post-test and the gains on the post-test were retained. Similarly, the Assisted RR group also showed statistically significant gains (i.e., 2.5% increase) in incidental word knowledge after the treatment; however, the gains in scores did not sustain on the delayed post-test. For both groups, a majority of the gains in scores were attributed to the gains in scores for the task that measured contextual recognition of words (i.e., the ability to provide L1 meaning of the L2 target word form). For the Control group, there was no statistically significant difference between the pre-test, the post-test, and the delayed post-test scores. (See Table 19 and Figure 8 in Chapter IV.) Additionally, there was no statistically significant difference between the groups for

the vocabulary pre-test or delayed post-test. A statistically significant difference was only found between the Unassisted RR group and the Control group on the post-test. As the participants' journal reflections indicated, many participants in the Unassisted RR group commented that they gradually understood the unknown words through repeated reading and that the words became easier. On the other hand, only one participant in the Assisted RR group expressed that the words became easier through repeated reading. In sum, the treatment led to statistically significant incidental vocabulary gains and retention for the Unassisted RR group and some amount of vocabulary gains for the Assisted RR group.

An explanation to the gains in scores from the pre-test to the post-test is that the study involved multiple encounters of the text and words and multiple treatment sessions, and thus provided essential conditions for acquisition to take place. Exposure to the same text multiple times within a session and to consecutive passages of an entire story throughout the sessions increases content familiarity of the text (Han & D'Angelo, 2009). Having such increased familiarity with the text may have provided additional contextual support for the participants such that the contextual redundancy provides assistance for the learners' acquisition of the word. Also, through frequent and repeated exposure to the text, the recurrence of vocabulary and linguistic items may have increased their saliency and resulted in an increased chance for them to be attended to and processed alongside meaning – which according to the SLA literature is an essential criterion for L2 acquisition in general (Han & D'Angelo, 2009; Schmidt, 2001), and therefore for vocabulary acquisition in particular. However, it is worthy of note that the participants were exposed to the target words during the pre-test. This may have heightened their awareness towards the target words during the reading process at the later sessions.

Therefore, it may not be just RR that impacted the result; the participants' exposure to the target words during pre-test may have played a contributing role in their learning of the target words.

In line with previous RR-L2 studies (e.g., Han & Chen, 2010; Liu & Todd, 2014, 2016; Webb & Chang, 2012; Zahar et al., 2001) that supported that RR leads to incidental vocabulary acquisition, the present study also showed statistically significant but differential gains in vocabulary scores as a result of RR. In Zahar et al. (2001) and Webb and Chang (2012), the participants were asked to subvocalize the sound code for the visual input, and the results showed 7%-13% retention of novel words. In Han and Chen (2010), the participant performed time-lapse imitation of the audio model and was able to retain the orthographic form for 55% of the unknown words. Compared to the other RR-L2 studies, Han and Chen's study yielded higher gains in scores, which most likely can be attributed to the provision of other form-oriented instructional strategies (including feedback and production tasks) to repeated reading, which were not present in the other studies. The provision of other instructional strategies may have called for deeper 'depth of information processing' (Craik & Lockhart, 1972). Research has generally suggested that reading or RR supplemented with various types of form-oriented instruction leads to better vocabulary acquisition and an increase in depth of word knowledge. Such additional resources were not available in the present study. Thus, the different treatment designs (e.g., provision of additional instructions) may have led to differences in depth of processing, and therefore to differential gains for the studies. Yet without providing additional resources, the present study still yielded positive results in the participants'

vocabulary acquisition, which suggests that RR used in isolation may still lead to incidental vocabulary acquisition.

Similar to the current study, other instructional strategies or resources were not included in Zahar et al. (2001) and Liu and Todd (2014, 2016). In their studies, multiple-choice word definition tests were used to measure vocabulary acquisition, and the participants encountered the words an average of seven times. Also, vocabulary was measured immediately or the day after exposure to the words, and the studies used graded readers and level-appropriate texts, respectively. On the other hand, the current study asked the participants to provide the word form or meaning, which is more difficult than a multiple-choice test. Additionally, the majority of words were encountered five times throughout the study, which is less than the seven encounters in Zahar et al. (2001) and Liu and Todd (2014, 2016). The higher number of encounters in the previous studies may have increased the saliency of the words and the participants' depth of processing. Furthermore, in the current study, the target words were not measured immediately and a more difficult text was used. Therefore, the methodological differences may have accounted for the higher scores in Liu and Todd's (2014, 2016) studies, compared to the present study. As far as vocabulary gains are concerned, the result of the present study was nonetheless compelling in that the participants were provided with more challenging conditions and less additional resources compared to the previous studies, yet under such circumstances, the present study still yielded promising results that were similar to Zahar et al.'s (2001) study. Furthermore, this underscores the need for more studies that systematically examine the contributions made by different components of RR (e.g.,

instruments, number of repetitions, the text difficulty level, the type and amount of external assistance, and the number of sessions needed).

Nonetheless, it is worth mentioning that none of the aforementioned RR-L2 studies on vocabulary acquisition incorporated the use of a delayed post-test in their design. The present study incorporated a delayed post-test, and both of the treatment groups showed some amount of retention of the gains and a higher score for the delayed post-test compared to the pre-test, although a statistically significant difference was found only for the Unassisted RR group. That the incidental vocabulary gains on the post-test were retained two months later for the Unassisted RR group is encouraging. Survey responses confirmed that none of the participants had much contact with English language learning between the post-test and the delayed post-test, nor did they study anything related to the present study outside of the sessions. This suggests that the long-term learning effect of the target words for the Unassisted RR group would have likely resulted from the treatment. It also suggests a need for studies to incorporate a delayed post-test in order to document whether vocabulary acquisition through RR retained.

Research into dual modality theory generally suggested that dual modality input leads to better learning outcomes than single modality input. Liu and Todd (2014), on the other hand, have pointed out that this may not always be the case. The beneficial effect of dual modality input varies depending on the learning condition and may not always be as clear-cut. In the current study, the benefit of dual modality input did not seem to be as promising as the literature suggests. The Unassisted RR group (single modality input) performed better than the Assisted RR group (dual modality input) on the vocabulary post-test and delayed post-test. Since the Unassisted RR group participants were able to

read at their own pace, they were given opportunities to process the many unknown words (including the target words) in the text. As some participants reflected in their learning journal, they used the words before and after the unknown word to guess its meaning, which may have led to better vocabulary acquisition. In contrast, the Assisted RR group participants were reading along with the audio model and could not stop to do so when encountering unknown words. In conclusion, the ability to use metacognitive strategies, as facilitated by the unassisted nature of the experimental condition, may have also contributed to the higher vocabulary gains and retention for the Unassisted RR group.

With respect to previous RR-L2 research, of particular interest for the present study are Liu and Todd (2014) and Webb and Chang (2012), which are the only studies to date that directly compared dual modality RR and single modality RR conditions in terms of vocabulary acquisition. In Liu and Todd's (2014) study, no significant difference was found between the *subvocalization* group (dual modality; silent reading + listening) and the *visual only* group (single modality; silent reading + uttering a nonword sound in their mouths every three to four seconds). Webb and Chang's (2012) study found that the assisted RR group out-performed the unassisted RR group on the incidental vocabulary acquisition post-test. In their study, the participants were provided with level-appropriate texts and additional resources (i.e., dictionary, raising of questions, peer discussion), which may have facilitated vocabulary acquisition. In contrast to Webb and Chang (2012), the present study provided the participants with a more difficult text and no access to additional resources. This suggests that text difficulty and the provision of additional resources may have played a pivotal role in the acquisition of vocabulary through repeated reading.

5.3 Reading Comprehension and Vocabulary Acquisition

A comparison between the groups revealed a trade-off pattern between comprehension and acquisition performance: Overall, the results showed an improvement for the vocabulary acquisition scores but not for the reading comprehension scores. While the Unassisted RR group showed higher gains on the vocabulary acquisition scores from the pre-test to the post-test, there was also slightly more decrease on the reading comprehension scores. On the other hand, the Assisted RR group showed a slightly better performance on the reading comprehension scores but less improvement on the vocabulary acquisition scores, compared to the Unassisted RR group (see Figure 4 and Figure 8).

These results raise the question of the amount of cognitive resources that are available for the participants to use for lower-level and higher-level processes – as well as the amount of resources that are freed up for the participants to re-allocate their attention to processing for meaning and/or processing for form. As research suggests, humans have limited cognitive capacity; thus, if the learners are caught up with lower-level processing, fewer resources will be available for higher-level comprehension processing (LaBerge & Samuels, 1974).

As in the present study, when the text is difficult – in terms of the participants' higher percentage of unknown words in the text (i.e., 10% unknown word density) – the participants may have experienced increased cognitive burden, because they may have struggled with lower-level processing. When L2 readers are cognitively over-loaded with lower-level processing and read with slow fluency, they lack sufficient resources for processing the text for higher-level comprehension, thus resulting in poor performance in

comprehension scores. Thus, the results of the present study suggest that a certain threshold (e.g., percentage of known words of a text) may need to be met before the beneficial effects of repeated reading can transpire for comprehension.

The results of the present study showed that provided with more challenging texts and no additional support, the participants were still able to benefit from the incidental vocabulary acquisition (i.e., processing the text for form) through repeated reading. SLA research has suggested that when the task is challenging, learners tend to pay more attention to form. As suggested by findings on input processing during L2 reading, L2 learners "need to attend to form for linguistic development" (Han & D'Angelo, 2009, p. 185). As a result of the participants paying more attention to form, vocabulary acquisition can become more likely. This may serve as a reason why the participants were able to process the text for form, and therefore achieve gains on the vocabulary acquisition test.

As discussed in Chapter III, the participants' EFL curriculum was heavily structure-oriented; this may also have affected the participants' processing orientation and resulted in better performance on the acquisition scores, as opposed to their comprehension test performance. As Han and Peverly (2007) suggested, instructed learners tend to engage in form-oriented processing, rather than the widely-adopted notion of meaning primacy (VanPatten, 1996, 2004) for learning. Since the participants' curriculum focused on forms, the learners may have been pre-programmed to constantly focus their attention on the formal features of the input. As a result of the potentially preexisting curriculum, the participants may have focused heavily on form-oriented processing and thus, the participants only benefited from vocabulary growth.

All in all, the fact that the participants were exposed to difficult texts, and that their EFL curriculum focused on forms, may have together created less favorable conditions for meaning-based comprehension processing and a more favorable condition for acquisition/processing of form. As Han (2007) noted, "simultaneously focusing learners' attention on meaning and form during input processing may lead to a trade-off effect, such that one is processed at the expense of the other … If these findings are confirmed by further research, *double treatment* should not be recommended for implementation in a pedagogical setting" (p. 390). The present study findings suggest that at the 90% known word density threshold and with no additional support, the trade-off effect favors vocabulary acquisition processing over reading comprehension processing. A call is made for more research that examines comprehension and acquisition in tandem so that we can come closer to a conclusion as to whether double treatment is effective for L2 development, or if there is a trade-off between comprehension and acquisition. Specifically, a next step in research would be to investigate the threshold of known word density through RR for reading comprehension and vocabulary acquisition for L2 learners. Only then can we suggest under which conditions double treatment should or should not be promoted.

5.4 Limitations and Future Directions

The present study had several notable limitations, which pertain to the effects of the study design on the entire study, the participants' comprehension test performance, and vocabulary test scores.

A first limitation pertains to the effects of the study design on the entire study. Although the participants read the same passage five times, the study cannot ensure that they were actually engaged in reading each time. The current study attempted to control for this by having the participants click through the screen five times (with a 30-second break in between each reading). Also, they were asked to rate their level of on-task-ness for each reading, and the majority of the participants that were included for data analysis provided a score of three or above (on a 5-point Likert scale) for each reading, indicating that they were on-task to being really on-task. However, this is a subjective measure and can only serve as an indirect support. Without a direct measure of where the participants' attention was focused during reading, it is impossible to provide direct and objective evidence whether they were actually engaged in reading each time. This is a limitation of the study and should be noted. Future studies would benefit from, for example, including another instrument that gauges the participants' thinking process (e.g., stimulated recall) to guarantee that the participants were actually reading and to get at internal validity.

As the present study findings and existing research suggested, text difficulty (as operationalized through known-word density, which is a learner internal measure) can modulate the potential benefits of repeated reading. Thus, future studies may benefit from administering vocabulary levels tests to the participants prior to the pre-test, rather than obtaining the information from other sources (e.g., pilot studies on other students from the same grade level and school), in order to really know the participants' level and forecast their comprehension of a text. Furthermore, doing so allows studies to gauge the effects of different text difficulty levels on the learners reading comprehension and vocabulary acquisition.

While the present study aimed to improve ecological validity by conducting research in school-based settings, the design of the study was nonetheless limited in many aspects. First, the study was limited in terms of the number of sessions available. Second is the impact of the number of remaining participants due to the participant elimination process (violation of the rules, etc.) and due to attrition. (See the Data Analysis section for details regarding dropped cases and attrition.) The elimination process resulted in a decrease in the number of participants remaining for data analysis. With a smaller number of participants, each participant may have had a larger impact on the result; thus, the results should be interpreted with caution. Due to the longer duration of the study, several participants could not come to all sessions. The attrition rate is a limitation of the study and should be noted and kept in mind for future research. Additionally, the post-test sessions took place toward the end of a semester with the final exam period approaching, and the participants may have been fatigued from preparing for their final exams. The participants' energy level may have affected their post-test performance and should be noted.

Furthermore, a possible limitation can be explained by the lack of random sampling of the participants. Due to practicality reasons, the study was limited to grouping participants according to their intact classes. Even though there was a similar number of participants in each group and the ANOVA results suggested that there was no statistically significant difference between the groups on the pre-test scores, future studies whenever possible should aim for random sampling of the participants. Alternatively, it may be more ideal to have a smaller group of participants come in for the experiment at a time – in lieu of a larger group working on the experiment at the same time – so as to

better monitor the participants' performance and control for unexpected behaviors such as violating the rules.

Certain limitations of the study are related to the effects of the study design on the participants' reading comprehension. First, a serious limitation is related to the comparability of the testing passages and its potential effect on the comprehension test results. As mentioned in the previous section, despite other factors being controlled for the passages, a post hoc analysis suggested that the post-test passage was slightly more difficult than the pre-test passage and the delayed post-test passage in reading demand of the text in terms of semantic difficulty and syntactic complexity, as well as the constantly changing of the settings. As mentioned in Chapter II, authentic input (i.e., text) contains features required for acquisition to occur and is in line with the communicative approach which aims at fostering the learners' ability to work with authentic texts (Han & D'Angelo, 2009); thus, the present study utilized an authentic text as reading material. Future studies utilizing original, authentic texts may benefit by, for example, making minimal modifications to the text so as to ensure comparability of the passages in every aspect.

Another limitation pertains to the role of the researcher in the coding process for the testing passages and the participants' recall protocols. As mentioned in the Data Coding section, due to schedule conflict, the norming sessions and the post-coding discussion sessions were carried out in different small groups, in lieu of having all coders meet at the same time. The researcher attempted to serve as a facilitator during the sessions. However, the researcher's subjectivity may have most likely had inadvertently influenced the results. Also, the researcher may have had an influence on decision-

making when discrepancy occurred. The role that the researcher played in the coding and decision making process for the testing passages and the participants' recall protocols is a limitation of the study and should be noted. Future studies should aim to conduct the norming and post-coding discussion sessions with all raters at the same time.

A third limitation is related to the two possible meanings behind a score of zero on the recall protocol. Across all testing sessions, three participants indicated that they forgot the content of the reading passage ("我忘了 [*I forgot*]"), whereas 18 participants reported that they could not understand the passage ("看不懂 [*I don't understand*]"). Even though these two types of answers were both given a score of "0," the scores suggest different underlying causes. Also, the zero score could imply that the participants have partial knowledge but did not understand the text to the point where they were able to access the content from memory, or that they understood some parts of the passage but had forgotten the content. Thus, the participants' actual comprehension scores could have been higher than the picture presented. Although only three participants indicated that they had forgotten the content of the reading passage, future studies would benefit from having alternative measures. For instance, future studies could check with the participants afterwards whether they had forgotten or they truly had no understanding of the text. Alternatively, researchers could provide the participants some training for the recall test task ahead of time before administering the actual recall test. Furthermore, future studies could consider including another type of comprehension measure; for instance, one that captures "emerging or developing or partial" knowledge. The participants can be provided with some sort of scaffolding through a semi-structured task, for example, a cued written recall.

Certain limitations of the study are related to the effects of the study design on the participants' vocabulary acquisition. First, a potential limitation of the study derived from the use of a pre-test. While the inclusion of the pre-test allows the study to document any developmental changes before and after the treatment, in doing so the participants were exposed to the target words in the pre-test, which may have heightened their awareness towards the target words during the reading process at the later sessions. The study attempted to reduce this potential threat by having a one-week intermission between the pre-test sessions and the start of the treatment period. Nevertheless, for future studies, a longer gap between vocabulary pre-test and reading of the passages may be needed.

A second limitation is the selection of the distractor words. Distractor words were selected from the reading materials. In lieu of using words from the text, the distractor words should have been selected from outside the reading material. Thus, the word "distractor" in the present study may not be a good fit and should be noted as a limitation.

Lastly, a limitation may have been caused by having the participants circle unknown words. In order to gather information on the participants' known word density, they were asked to circle unknown words prior to reading each passage. It is worth noting that performance of this task may have heightened their awareness towards their own unknown words and may have reoriented the learners' attention resources toward form-based processing. Future studies can consider having the participants identify unknown words for the following passage at the end of each session. Alternatively, future studies could benefit from having the participants identify their unknown words for all the passages a few weeks prior to the pre-test, if possible. In doing so the study can obtain a baseline of the participants' unknown word density of the whole story.

5.5 Conclusion

To conclude, this study set out to examine whether L2 repeated reading –

unassisted and assisted – promotes reading comprehension development and leads to

incidental vocabulary acquisition. The findings suggest that L2 repeated reading did not

promote reading comprehension and that it did not contribute to a transfer of practice

effect to new text in terms of reading comprehension. However, there were statistically

significant incidental vocabulary gains and retention for the Unassisted RR group and

some amount of vocabulary gains for the Assisted RR group. When the unknown word

density reached 10% and the participants were not provided with additional support, five

repeated encounters with the text (over eight treatment sessions) were found to be

inadequate in promoting reading comprehension. Nonetheless, the participants provided

with such challenging condition still benefited from the incidental vocabulary acquisition.

This may imply that a certain threshold of proficiency (e.g., percentage of known words

of a text) is necessary for the beneficial effects of repeated reading to support

comprehension. Additionally, repeated reading alone may still be insufficient for EFL

learners dealing with difficult texts and that additional support to repeated reading may

still be necessary. All in all, the fact that the participants were exposed to difficult texts

may have created less favorable conditions for meaning-based comprehension processing

and a more favorable condition for acquisition/processing of form.

This study provides a counterpoint to previous findings in the literature regarding

the effects of L2 repeated reading on reading comprehension and on facilitating reading

of difficult text. On the other hand, the study reinforces previous findings in the literature

that L2 repeated reading leads to incidental vocabulary acquisition. The effects of dual

and single modality input in L2 repeated reading remain inconclusive and motivate ongoing study in this strand of L2 reading research. The seemingly complex findings of this study suggest that the relationship between comprehension and acquisition, as far as the dual goals of reading are concerned for L2 learners, is not straightforward. Thus, more research is needed to address the efficacy of the repeated reading approach within the dual relevance of L2 reading framework so that we can come closer to a conclusion as the effects of L2 repeated reading on comprehension and acquisition and under which conditions double treatment should or should not be promoted.

REFERENCES

Aebersold, J., & Field, M. (1997). *From reader to reading teacher*. Cambridge, UK: Cambridge University Press.

Ahmed, S., & Han, Z-H. (in press). Models of reading. In J. I. Liontas (Ed.), TESOL International Association, *The TESOL encyclopedia of English language teaching*. Oxford, UK: Wiley-Blackwell.

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

Anderson, N. (1993). Repeated reading. In R. R. Day (Ed.), *New ways in teaching reading* (pp. 190–191). Alexandria, VA: Teachers of English to Speakers of Other Languages.

Anderson, N. (2009). ACTIVE reading: The research base for a pedagogical approach in the reading classroom. In Z.-H. Han & N. Anderson (Eds.), *L2 reading research and instruction: Crossing the boundaries* (pp. 117–143). Ann Arbor, MI: The University of Michigan Press.

Anderson, J. R., & Bower, G. H. (2013). *Human associative memory*. New York, NY: Psychology Press.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading/Language Research, 2*, 231-256.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*, 189–208.

Baddeley, A. D. (2006). Working memory: An overview. In Pickering S. (Ed.), *Working memory and education* (pp. 1-31). Burlington, MA: Academic Press.

Baddeley, A. D. (2007). *Working memory, thought and action*. New York, NY: Oxford University Press.

Berkemeyer, V. C. (1989). Qualitative analysis of immediate recall protocol data: Some classroom implications. *Die Unterrichtspraxis, 21*(3), 131-137.

Bernhardt, E. B. (1983a). Testing foreign language reading comprehension: The immediate recall protocol. *Die Unterrichtspraxis, 16*(1)*, 27-33.*

Bernhardt, E. B. (1983b). Three approaches to reading comprehension in intermediate German. *Modern Language Journal, 67*(2)*,* 111-115.

Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical, and classroom perspectives*. Norwood, NJ: Ablex Publishing Corporation.

Bernhardt, E. B. (1992). A psycholinguistic perspective on second language literacy. *AILA Review*, 8, 31-44.

Bernhardt, E. B. (2011). *Understanding advanced second-language reading*. New York, NY: Routledge.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. G. A. J. Schachter (Ed.), *Linguistic perspectives on second language acquisition* (pp. 41-68). Cambridge, UK: Cambridge University Press.

Blum, I. H., Koskinen, P. S., Tennant, N., Parker, E. M., Straub, M., & Curry, C. (1995). Using audiotaped books to extend classroom literacy instruction into the homes of second-language learners. *Reading Research Report, 39*, 1-32.

Brantmeier, C. (2006). The effects of language assessment and L2 reading performance on advanced readers' recall. *The Reading Matrix, 6*, 1-17.

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

Carrell, P. L. (1983). Some issues in studying the role of schemata, or background knowledge, in second language comprehension. *Reading in a Foreign Language, 1*, 81-92.

Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly, 19*(4), 727-752.

Carver, R. (1990). *Reading rate: A review of research and theory*. San Diego, CA: Academic Press.

Carver, R. P., & Hoffman, J. V. (1981). The effect of practice through repeated reading on gain in reading ability using a computer-based instructional system. *Reading Research Quarterly, 16,* 374-390.

Chang, C.-S. (2012). Improving reading rate activities for EFL students: Timed reading and repeated oral reading. *Reading in a Foreign Language, 22*, 43-62.

Chang, C.-S., & Millett, S. (2013). Improving reading rates and comprehension through timed repeated reading. *Reading in a Foreign Language, 25*(2), 126-148.

Chang, Y.-F. (2006). On the use of the immediate recall task as a measure of second language reading comprehension. *Language Testing, 23*(4), 520-543.

Chen, C.-L. A. (2013). Multilingualism and the holistic approach to multilingual education. *Working Papers in TESOL & Applied Linguistics, 13*(2), 45-46.

Chomsky, C. (1976a). After decoding: What? *Language Arts, 53,* 288-296.

Chomsky, C. (1976b). Approaching reading through invented spelling. [Presentation]. *Paper presented at the Conference on Theory and Practice of Beginning Reading Instruction.* Pittsburgh, PA.

Chomsky, C. (1978). When you still can't read in third grade. After decoding, what? In S. Samuels (Ed.), *What research has to say about reading instruction* (pp. 13-30). Newark, DE: International Research Association.

Cobb, T. *Compleat Lexical Tutor v.8* [computer program]. Accessed March 15, 2016 at https://www.lextutor.ca.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology, 11,* 38-63.

Cobb, T. (2009). Necessary or nice? Computers in second language reading. In Z.-H. Han & N. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 144-172). Ann Arbor, MI: University of Michigan Press.

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

Cromer, W. (1970). The difference model: A new explanation for some reading difficulties. *Journal of Educational Psychology, 61*, 471-483.

Dahl, P. J. (1974). *An experimental program for teaching high speed word recognition and comprehension skills.* Washington, DC: National Institute of Education, Office of Research. (ERIC Document Reproduction Service No. ED 099812)

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433.

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

Dlugosz, D. W. (2000). Rethinking the role of reading in teaching a foreign language to young learners. *ELT Journal, 54*(3), 284-290.

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly, 22,* 389-406.

Dowhower, S. L. (1989). Repeated reading: Research into practice. *The Reading Teacher, 42*(7), 502–507.

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

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*, 174-187.

Faulkner, H. J., & Levy, B. A. (1994). How text difficulty and reader skill interact to produce differential reliance on word and content overlap in reading transfer. *Journal of Experimental Child Psychology, 50,* 1-24.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in reading a second language. *Memory and Cognition, 11*, 565-574.

Fitzgerald, J., & Spiegel, D. L. (1983). Enhancing children's reading comprehension through instruction in narrative structure. *Journal of Reading Behavior, 15*(2), 1-17.

Fleisher, L. S., Jenkins, J. R., & Pany, D. (1979). Effects on poor readers' comprehension of training in rapid decoding. *Reading Research Quarterly, 15*, 30-48.

Fraser, C. (2007). Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *The Modern Language Journal, 91*, 372-394.

Gass, S., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gass, S. (with Behney, J., & Plonsky, L.) (2013). *Second language acquisition: An introductory course* (4th ed.). New York, NY: Routledge.

Gorsuch, G., & Taguchi, E. (2008). Repeated reading for developing reading fluency and reading comprehension: The case of EFL learners in Vietnam. *System, 36*, 253-278.

Gorsuch, G., & Taguchi, E. (2010). Developing reading fluency and comprehension using repeated reading: Evidence from longitudinal student reports. *Language Teaching Research, 14*, 27-59.

Gorsuch, G., Taguchi, E., & Umehara, H. (2015). Repeated reading for Japanese language learners: Effects on reading speed, comprehension, and comprehension strategies. *The Reading Matrix*, *15*(2), 18-44.

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

Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics, 24*, 44-69.

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

Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading*. New York, NY: Routledge.

Han, Z.-H. (2007). Pedagogical implications: Genuine or pretentious? *TESOL Quarterly, 41*(2), 387-393.

Han, Z.-H., Anderson, N. J., & Freeman, D. (2009). Introduction: Crossing the boundaries. In Z.-H. Han & N. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 1-13). Ann Arbor, MI: University of Michigan Press.

Han, Z.-H., & Chen, C.-L. A. (2010). Repeated-reading-based instructional strategy and vocabulary acquisition: A case study of a heritage speaker of Chinese. *Reading in a Foreign Language, 22*(2), 242-262.

Han, Z.-H., & D'Angelo, A. (2009). Balancing between comprehension and acquisition: Proposing a dual approach. In Z.-H. Han & N. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 173-191). Ann Arbor, MI: University of Michigan Press.

Han, Z. H, & Peverly, S. T. (2007). Input processing: A study of ab initio learners with multilingual backgrounds. *International Journal of Multiculturalism, 4*, 17-37.

Herman, P. A. (1985). The effect of repeated readings on reading rate, speech pauses, and word recognition accuracy. *Reading Research Quarterly, 20*, 553-564.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis 15*(3), 199-236.

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

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes, 33*, 54-65.

Hu, M. H.-C., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*, 403-430.

Hudson, T. (2007). *Teaching second language reading*. New York, NY: Oxford University Press.

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second language learning: Introduction. *Studies in Second Language Acquisition, 27*(2), 129-140.

Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica, 15*, 259-280.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160-212.

Johnston, P. H. (1983). *Reading comprehension assessment: A cognitive basis.* Newark, DE: International Reading Association.

Kintsch, W., Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working-memory in text comprehension. *Psychologica, 42*, 186-198.

Klingner, J. K. (2004). Assessing reading comprehension. *Assessment for Effective Intervention, 29*(4), 59-70.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach.* New York, NY: Cambridge University Press.

Koskinen, P. S., & Blum, I. H. (1984). Repeated oral reading and the acquisition of fluency. In J. Niles & L. Harris (Eds.), *Changing perspectives on research in reading/language processing and instruction.* Thirty-third Yearbook of the National Reading Conference. National Reading Conference, Rochester, NY, pp. 183-187.

Krashen, S. (1985). *The input hypothesis: Issues and implications.* London, UK: Longman.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95,* 3-21.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323.

Laerd Statistics (2016). Two-way mixed ANOVA using SPSS Statistics. *Statistical Tutorials and Software Guides.* Retrieved from *https://statistics.laerd.com/.*

Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review, 63*(1), 149-166.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15.

Lee, J. F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition, 8,* 201-212.

Lee, J. F. (1998). The relation of verb morphology to second language reading comprehension and input processing. *The Modern Language Journal, 82*(1), 33-48.

Lee, J. F. (2002). The initial impact of reading as input for the acquisition of future tense morphology in Spanish. In S. Gass, K. Bardovi-Harloig, S. Magnan, & J. Walz (Eds.), *Pedagogical norms for second and foreign language learning and teaching* (pp. 119-140). Amsterdam, Netherlands: John Benjamins.

Lee, G., & Young, V. (1974). Auditory versus visual: A dual modality theory. *Kappa Delta Pi Record, 11*, 28-30.

Leow, R., Hsieh, H., & Moreno, N. (2008). Attention to form and meaning revisited. *Language Learning, 58*(3), 665-695.

Liao, Y.-F. (2009). *A construct validation study of the GEPT reading and listening sections: Re-examining the models of L2 reading and listening abilities and their relations to lexico-grammatical knowledge.* (Ed.D. dissertation), Teachers College, Columbia University.

Lightbown, P. M. & Spada, N. (1990). Focus on form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition, 12*(4), 429-448.

Liu, Y.-T., & Todd, A. G. (2014). Dual-modality input in repeated reading for foreign language learners with different learning styles. *Foreign Language Annals, 47*(4), 684-706.

Liu, Y.-T., & Todd, A. G. (2016). Implementation of assisted repeated reading techniques for the incidental acquisition of novel foreign vocabulary. *Language Teaching Research, 20*(1), p. 53-74.

Logan, G. (1997). Automaticity and reading: Perspective from the instance theory of automatization. *Reading and Writing Quarterly*, *13*, 123–146.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, and C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam, Netherlands: John Benjamins.

Long, M. H. (2017). Instructed second language acquisition (iSLA): Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition, 1*(1), 7-44.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom language acquisition* (pp. 15-41). Cambridge, UK: Cambridge University Press.

Morgan, R., & Lyon, E. (1979). Paired reading: A preliminary report on a technique for parental intuition on reading to retarded children. *Journal of Child Psychology and Psychiatry, 20*, 151-160.

Moyer, S. B. (1982). Repeated reading. *Journal of Learning Disabilities, 15*(10), 619-623.

Nation, I. S. P. (2001). *Vocabulary learning in another language.* New York, NY: Cambridge University Press.

Nation, I. S. P. (2005). Reading faster. *PASAA, 36*, 21-35.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59-82.

National Center for Education Evaluation and Regional Assistance (2014). *WWC procedures and standards handbook, Version 3.0*. Institute of Education Sciences (IES), U.S. Department of Education. Retrieved from *https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_stan dards_handbook.pdf*.

National Reading Panel (2000). *Report of the subgroups: National Reading Panel.* Washington, DC: National Institute of Child Health and Human Development.

National Research Council (2012). *Improving adult literacy instruction: Options for practice and research*. Washington, D.C.: National Academies Press.

O'Shea, L. J., Sindelar, P. T., & O'Shea, D. J. (1985). The effects of repeated readings and attentional cues on reading fluency and comprehension. *Journal of Reading Behavior, 17*(2), 129-142.

Osa-Melero, L. (2015). A quantitative comparison of the effects of small group versus individual pre-reading activities on the reading comprehension of college students of Spanish. *Electronic Journal of Foreign Language Teaching, 12*(1).

Perfetti, C. A., & Lesgold, A. M. (1976). Discourse comprehension and sources of individual differences. In M. A. Just, & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp.141-184). Hillsdale, NJ: Erlbaum.

Perfetti, C. A., & Roth, S. (1981). Some of the interactive processes in reading and their role in reading skill. In A. Lesgold & C. Perfetti (Eds.), *Interactive processes in reading* (pp. 269–297). Hillsdale, NJ: Erlbaum Associates.

Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching* (3rd ed.). New York, NY: Guilford.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning 52*(3), 513–536

RAND Reading Study Group (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Retrieved October 14, 2014, from *www.rand.org/multi/achievementforall/reading/readreport.html.*

Rashotte, C. A., & Torgesen, J. K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly, 20,* 180-188.

Riley, G., & Lee, J. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing, 13*, 173-189.

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2016). Does modality matter? The effects of reading, listening, and dual modality on comprehension. *Sage Open, 6*(3), 1-9.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*, 589-619.

Samuels, S. J. (1976). Automatic decoding and reading comprehension. *Language Arts, 53*, 323-325.

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 32*(4), 403-408.

Samuels, S. J. (2006). Reading fluency: Its past, present, and future. In T. Rasinski, C. Blachowicz, & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (pp. 7-20). New York, NY: Guilford Press.
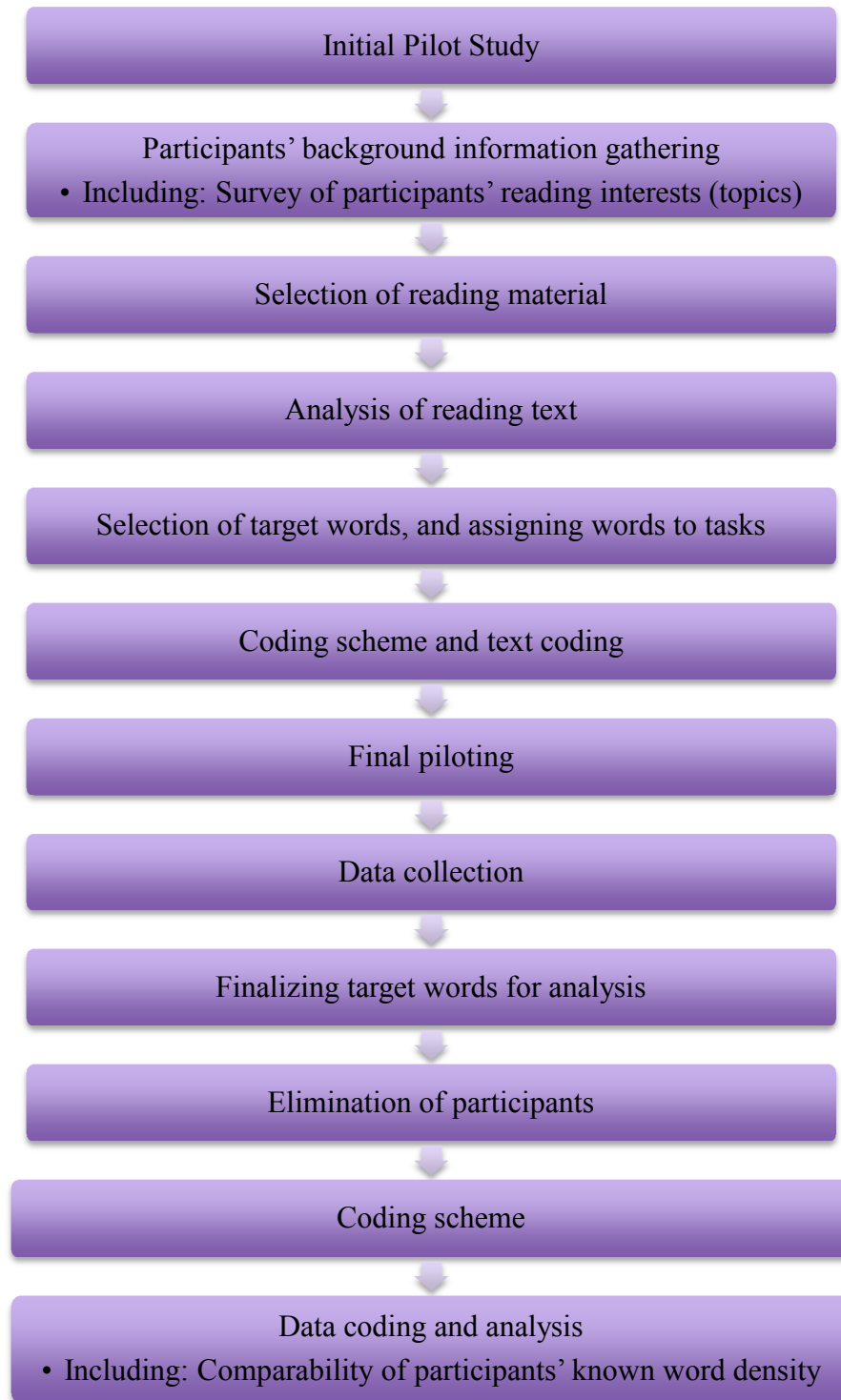
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching, 50*(2), 212-226.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comrpehension. *Modern Language Journal, 95*(1), 26-43.

Schreiber, P. A. (1980). On the acquisition of reading fluency. *Journal of Reading Behavior*, *12*, 177–186.

Segalowitz, N., Poulsen, C., & Komoda, M. (1991). Lower level components of reading skill in higher level bilinguals: Implications for reading instruction. *AILA Review, 8*(1), 15-30.

Sharwood Smith, M. (1986). Comprehension versus acquisition: Two ways of processing input. *Applied Linguistcs, 7*(3), 239-256.

Shiotsu, T. (2009). Reading ability and components of word recognition speed: The case of L1-Japanese EFL learners. In Z. -H. Han & N. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 15–39). Ann Arbor, MI: The University of Michigan Press.

Sindelar, P. T., Monda, L. E., & O'Shea, L. J. (1990). Effects of repeated readings on instructional- and mastery-level readers. *Journal of Educational Research, 83*, 220-226.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford, UK: Oxford University Press.

Smith, F. (1978) *Understanding reading: A psycholinguistic analysis of reading and learning to read*. New York, NY: Holt, Rinehart & Winston.

Taguchi, E. (1997). The effects of repeated readings on the development of lower identification skills of FL readers. *Reading in a Foreign Language, 11*(1), 97-119.

Taguchi, E., & Gorsuch, G. J. (2002). Transfer effects of repeated EFL reading on reading new passages: A preliminary investigation. *Reading in a Foreign Language, 14*(1), 43-65.

Taguchi, E., & Gorsuch, G. J. (2012). Fluency instruction in reading in a second or foreign language. In T. Rasinski, C. Blachowicz, & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (2nd ed.) (pp. 255-288). New York, NY: The Guilford Press.

Taguchi, E., Gorsuch, G. J., & Sasamoto, E. (2006). Developing second and foreign language reading fluency and its effect on comprehension: A missing link. *The Reading Matrix, 6*(2), 1-18.

Taguchi, E., Gorsuch, G., Lems, K., & Rosszell, R. (2016). Scaffolding in L2 reading: How repetition and an auditory model help readers. *Reading in a Foreign Language, 28*(1), 101-117.

Taguchi, E., Gorsuch, G., Takayasu-Maass, M., & Snipp, K. (2012). Assisted repeated reading with an advanced-level Japanese EFL reader: A longitudinal diary study. *Reading in a Foreign Language, 24*(1), 30-55.

Taguchi, E., Takayasu-Maass, M., & Gorsuch, G. J. (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.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25*(4), 252-261.

Thompson, I. (1995). Testing listening comprehension. *AATSEEL Newsletter, 37*, 24-31.

Thornbury, S. (2002). *How to teach vocabulary.* Harlow, UK: Pearson.

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System, 21*(3), 371-380.

Tucker, G. R. (1998). A global perspective on multilingualism and multilingual education. In J. Cenoz, & F. Genesee (Eds.), *Beyond bilingualism: Multilingualism and multilingual education* (pp. 3-15). Clevedon, UK: Multilingual Matters.

Valdes, G. (2012). *Multilingualism.* Retrieved October 31, 2013, from *www.linguisticsociety.org/resource/multilingualism.*

VanPatten, B. (1990). Attention to form and content in the input. *Studies in Second Language Acquisition, 12,* 287-301.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research.* Norwood, NJ: Albex.

VanPatten, B. (2004). Input Processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-31). Mahwah, NJ: Lawrence Erlbaum Associates.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*, 33-52.

Webb, S., & Chang, A. C-S. (2012). Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review, 68*(3), 267-290.

Wells, D. R. (1986). The assessment of foreign language reading comprehension: Refining the task. *Die Unterrichtspraxis, 19*(2), 178-184.

Young, A. R., Bowers, P. G., & MacKinnon, G. E. (1996). Effects of prosodic modeling and repeated reading on poor readers' fluency and comprehension. *Applied Psycholinguistics, 17,* 59-84.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness**.** *Canadian Modern Language Review*, *57,* 541–572.

Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. (1977). A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition. *Intelligence, 1*, 5-31.

Appendix A

Background Preparation Procedure

Initial Pilot Study

Participants' background information gathering
• Including: Survey of participants' reading interests (topics)

Selection of reading material

Analysis of reading text

Selection of target words, and assigning words to tasks

Coding scheme and text coding

Final piloting

Data collection

Finalizing target words for analysis

Elimination of participants

Coding scheme

Data coding and analysis
• Including: Comparability of participants' known word density

Appendix B

Background Questionnaire

(Original Version)

親愛的同學您好：

本問卷的目的在於了解您過去和現在的英語及閱讀學習狀況。

本問卷資料僅做為本「英語閱讀教學法」研究使用，不另作他用; 您的學號及填寫內容絕不對外透露。
您的意見對本研究及英語學習教育界有相當寶貴的貢獻，並且期望能對您的英文學習有所幫助。
衷心感謝您的參與及協助!

敬 祝
　　　　學 安
紐約哥倫比亞大學教育學院英語教學所
博士班研究生：陳貞伶 (Alice) 謹啟
指導教授：Dr. ZhaoHong Han

Q1. 您的答案不會影響到您的英文科成績，因此請您以最真實的答案作答。

若您願意參加此研究，請在此輸入您的學號：

Q2. 您的性別：

○ 男

○ 女

Q3. 您的年齡：

Q4. 您的母語為?
（可複選）

☐ 中文（國語．漢語）

☐ 台語

☐ 客語

☐ 英語

☐ 其它
　　[　　　　　]

Q5. 成長過程中, 您的家人會使用"哪(幾)種"語言與您溝通?
（可複選）

☐ 中文（國語．漢語）

☐ 台語

☐ 客語

☐ 英語

☐ 其它
　　[　　　　　]

Q6. 家裡 "最主要" 使用的語言為?
（可複選）

☐ 中文（國語．漢語）

☐ 台語

☐ 客語

☐ 英語

☐ 其它
　　[　　　　　]

Q7. 在不同的階段裡, 學校環境會使用的語言有哪些? （請簡述各種情況）

（例如: 國小: 中文
國中: 中文＆英文
高中/職: 中文＆英文）

[　　　　　　　　　　　　　　]

Q8. 您的生活中"最主要"使用的語言為?

[　　　　　　　　　　　　　　]

Qqualtrics

Q9. 請問您學英語已有多久的時間呢?
（幾年＋幾個月）

Q10. 請問您幾歲開始學英語呢?

Q11. 請問您是否曾經在英語系國家 短期或長期 居住過呢?

○ 是

○ 否

○ 其它

Q12. （承上題）若有, 大約幾歲到幾歲時?

（例如: 5~12歲 & 18~21歲）

Q13. 在成長過程中, 請問您有哪些英語課程學習經驗（例如: 補習班, 私人機構, 家教
班, 學校等）?
請簡述各種學習情況。

C. Timing

此網頁計時器將不會被顯示給接收者查看。
First Click: *0 seconds.*
Last Click: *0 seconds.*
Page Submit: *0 seconds.*
Click Count: *0 clicks.*

>>

Q１４. 除了學校的英文課之外，您過去六個月是否有參加任何英語相關課程（學校, 補習班, 私人機構, 家教班等）？

○ 是

○ 否

○ 其它

Q１５. （承上題） 若有, 已經上多久的課了？ 上課內容為何？
（請簡述各種學習情況。）

Q１６. 您過去六個月 "平均每週" 花多少時間上各式英語課程？ （校內 + 校外）

Q17. 您過去六個月，每週大約花多少時間 "自修" 英語？ （學校以外的場合：包含家裡 + 補習班）

Q１８. （承上題） 您的課外英語自修的內容為何？ （例如：看電影，上網，故事書，ＣＤ，報紙）

(沒有則填"無")

Q１９. 對英語學習之看法:

	非常不同意	不同意	中立	同意	非常同意
我對學習 "英文" 感到興趣。	○	○	○	○	○
我對學習英文 "閱讀" 感到興趣。	○	○	○	○	○
我對學習英文 "單字" 感到興趣。	○	○	○	○	○
我覺得英文單字很重要。	○	○	○	○	○

qualtrics

Q20.您的 "母語" 語言程度大致為何？（自我評量）

（1代表"較弱"之項目; 5代表"較強"之項目）

	1（較弱）	2	3	4	5（較強）
閱讀	○	○	○	○	○
聽力	○	○	○	○	○
單字／詞彙	○	○	○	○	○
寫作	○	○	○	○	○
會話	○	○	○	○	○
整體能力	○	○	○	○	○

Q21.您的 "英語" 語言程度大致為何？（自我評量）

（1代表"較弱"之項目; 5代表"較強"之項目）

	1（較弱）	2	3	4	5（較強）
閱讀	○	○	○	○	○
聽力	○	○	○	○	○
單字／詞彙	○	○	○	○	○
寫作	○	○	○	○	○
會話	○	○	○	○	○
整體能力	○	○	○	○	○

E. Timing

此網頁計時器將不會被顯示給接收者查看。

First Click: *0 seconds.*

Last Click: *0 seconds.*

Page Submit: *0 seconds.*

Click Count: *0 clicks.*

>>

Q qualtrics

Q22. 您是否曾經考過下列之英語能力檢定呢?
（若有，請勾選所有曾經考過之檢定考。）

☐ 全民英檢 (GEPT) - 初級

☐ 全民英檢 (GEPT) - 中級

☐ 全民英檢 (GEPT) - 中高級

☐ 多益 (TOEIC)

☐ 雅思 (IELTS)

☐ 無

☐ 其它

Q23. 請列舉您所參加過的英語檢定考之：(1) 總成績; (2) 閱讀分數 (若有); (3) 單字/字彙成績 (若有)。

若無, 請回答"無"。

Q24. 您未來六個月是否有計劃要考任何英語檢定考試?

○ 是 （檢定考名稱為? ）

○ 否

F. Timing

此網頁計時器將不會被顯示給接收者查看。
First Click: *0 seconds.*
Last Click: *0 seconds.*
Page Submit: *0 seconds.*
Click Count: *0 clicks.*

>>

0%   100%

Q25. 在 "母語" 學習過程中, 是否曾經有人教導您 "閱讀方法與技巧"?

○ 是 （請簡述之）

○ 否

Q26. 在 "英語" 的學習過程, 是否曾經有人教導您 "英語 閱讀方法與技巧"?

○ 是 （請簡述之）

○ 否

Q27. 在閱讀英文文本 （如: 英文課本, 小說, 等） 時, 您會使用哪些閱讀方法與技巧?

Q28. 試回想: 在單字學習上, 您通常使用哪些 "單字學習技巧" 呢?

（可複選）

☐ 默唸背誦
☐ 死記
☐ 字卡
☐ 從整篇文章的上下文猜字義
☐ 看單字在句子中怎麼用 （例句）
☐ 查字典
☐ 其它 （請說明）

Q29. 請問您遇到不會的英文單字, 您會如何處理?

（可複選）

☐ 問同學
☐ 問老師
☐ 查字典
☐ 猜猜看什麼意思
☐ 不管它
☐ 其它 （請說明）

Q30. 您是否喜歡閱讀中文小說 和 英文小說?

	很不喜歡	不喜歡	中立	喜歡	非常喜歡
中文小說	○	○	○	○	○
英文小說	○	○	○	○	○

(English Translation Version)

Q1.
This is a reminder that your answers will be used for this study only and your information will remain confidential. Thus, please provide your truthful answers.

If you agree to participate in this study, please write your student ID number here:

Q2. Your gender:

○ Male

○ Female

Q3. Your age:

Q4. What is/are your native language(s)?
(You may select more than one answer.)

☐ Mandarin Chinese

☐ Taiwanese

☐ Hakka

☐ English

☐ Other

Q5. What language(s) does/do your family use at home for communication?
(You may select more than one answer.)

☐ Mandarin Chinese

☐ Taiwanese

☐ Hakka

☐ English

☐ Others

[   ]

Q6. **What is/are the most commonly used language(s) at home?**
(You may select more than one answer.)

☐ Mandarin Chinese

☐ Taiwanese

☐ Hakka

☐ English

☐ Other

[   ]

Q7.

What is/are the language(s) used at school?
(Please briefly describe the situation for each education stage.)

(For example:
Elementary school: Mandarin Chinese
Junior high school: Chinese & English
High school: Chinese & English)

[   ]

Q 8 . What is the "most" commonly used language(s) for your current daily life?

Q 9 . How long have you studied English?
(year + month)

Q 1 0 . At what age did you start learning English?

Q 1 1 . Have you ever lived (short or long term) in an English-speaking environment?

○ Yes

○ No

○ Other

Q 1 2 . If so, when (what ages) did you live in an English-speaking environment?

(For example: 5~12 years old & 18~21 years old)

Q 1 3.
What settings (e.g., night/cram schools, private institution, tutoring, school settings) have you taken English lessons?

Please briefly describe each situation.

Q 1 4. In the past 6 months, have you attended English classes (e.g., night/cram school, private institution, tutoring) outside of school?

○ Yes

○ No

○ Other

Q 1 5. If so, how long have you taken the classes? And what content does the classes cover?

(Please briefly describe each situation.)

Q 1 6. In the past 6 months, how many hours a week (on average) did you spend on attending English classes? (both at school and after school)

Q17. In the past 6 months, how many hours a week did you spend on studying English on your own? (Outside of school settings: including at home and English lessons attended)

Q1 8. What did you use to learn English on your own? (For example: watch movies, surf the Internet, read story books, listen to songs, listen to CDs, read the newspaper, etc.)

(Please write "N/A" if this does not apply to your situation.)

Q1 9. What are your impressions for the following statements:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I'm interested in learning English.	○	○	○	○	○
I'm interested in learning English "reading".	○	○	○	○	○
I'm interested in learning English "vocabulary"	○	○	○	○	○
I think English vocabulary is very important.	○	○	○	○	○

Q 2 0 . How would you rate your native-language proficiency?

(1 means "very weak"; 5 means "very strong")

	1 (Very weak)	2	3	4	5 (Very strong)
Reading	○	○	○	○	○
Listening	○	○	○	○	○
Vocabulary	○	○	○	○	○
Writing	○	○	○	○	○
Speaking	○	○	○	○	○
Overall	○	○	○	○	○

Q 2 1 . How would you rate your "English" language proficiency?

(1 means "very weak"; 5 means "very strong")

	1 (Very weak)	2	3	4	5 (Very strong)
Reading	○	○	○	○	○
Listening	○	○	○	○	○
Vocabulary	○	○	○	○	○
Writing	○	○	○	○	○
Speaking	○	○	○	○	○
Overall	○	○	○	○	○

Q2 2. Have you taken any standardized English language proficiency assessment listed below?
(If so, please check off the test(s) you have taken before.)

- ☐ GEPT - beginner level

- ☐ GEPT - intermediate level

- ☐ GEPT - high intermediate level

- ☐ TOEIC

- ☐ IELTS

- ☐ None of the above

- ☐ Other

Q2 3. If possible, please provide the scores for the tests you have taken before: (1) Total score, (2) reading score (if available); (3) vocabulary section score (if available).

(Please write "N/A" if this is not applicable.)

Q2 4. Are you planning on taking any standardized English language proficiency assessment in the following six months?

- ○ Yes. (Please provide the title of the test you are planning on taking.)

- ○ No.

Q 2 5 . In acquiring your mother tongue, were you ever taught reading skills and strategies?

○ Yes. (Please briefly describe the situation.)

○ No.

Q 2 6 . In the process of learning English, were you ever taught English reading skills and strategies?

○ Yes. (Please briefly describe the situation.)

○ No.

Q 2 7 . When reading English passages (e.g., textbooks, novels, etc.), what reading skills and strategies do you use?

Q 2 8 . When learning a new English word, what strategies do you usually use?

(You may choose more than one answer.)

☐ Rehearse in my head until I remember the word

☐ Rote learning

☐ Use of vocabulary flashcards

☐ Use the context to guess its meaning

☐ Look at how the word is used in a sentence

☐ Look up the dictionary

☐ Other (Please describe)

Q 2 9 . What do you do when you encounter unknown English words?

(You may choose more than one answer.)

- ☐ Ask my classmates
- ☐ Ask my teacher
- ☐ Look up the dictionary
- ☐ Guess its meaning
- ☐ Don't do anything about it
- ☐ Other (Please describe)

Q 3 0 . Do you like to read Chinese novels or English novels?

	Strongly dislike	Dislike	Neutral	Like	Strongly like
Chinese novels	○	○	○	○	○
English novels	○	○	○	○	○

31. Lastly, do you have any suggestions or additional comments for this questionnaire?

Thank you!

Appendix C

Reading Materials

Testing: Pre-test

The Tale of Samuel Whiskers

ONCE upon a time there was an old cat, called Mrs. Tabitha Twitchit, who was an anxious parent. She used to lose her kittens continually, and whenever they were lost they were always in mischief!

On baking day she determined to shut them up in a cupboard.

She caught Moppet and Mittens, but she could not find Tom.

Mrs. Tabitha went up and down all over the house, mewing for Tom Kitten. She looked in the pantry under the staircase, and she searched the best spare bedroom that was all covered up with dust sheets. She went right upstairs and looked into the attics, but she could not find him anywhere.

It was an old, old house, full of cupboards and passages. Some of the walls were four feet thick, and there used to be queer noises inside them, as if there might be a little secret staircase. Certainly there were odd little jagged doorways in the wainscot, and things disappeared at night—especially cheese and bacon.

Mrs. Tabitha became more and more distracted, and mewed dreadfully.

While their mother was searching the house, Moppet and Mittens had got into mischief.

The cupboard door was not locked, so they pushed it open and came out.

They went straight to the dough which was set to rise in a pan before the fire.

They patted it with their little soft paws — "Shall we make dear little muffins?" said Mittens to Moppet.

Treatment: Session 1

But just at that moment somebody knocked at the front door, and Moppet jumped into the flour barrel in a fright.

Mittens ran away to the dairy, and hid in an empty jar on the stone shelf where the milk

pans stand.

The visitor was a neighbor, Mrs. Ribby. She had called to borrow some yeast.

Mrs. Tabitha came downstairs mewing dreadfully — "Come in, Cousin Ribby, come in, and sit ye down! I'm in sad trouble, Cousin Ribby," said Tabitha, shedding tears. "I've lost my dear son Thomas. I'm afraid the rats have got him." She wiped her eyes with an apron.

"He's a bad kitten, Cousin Tabitha. He made a cat's cradle of my best bonnet last time I came to tea. Where have you looked for him?"

"All over the house! The rats are too many for me. What a thing it is to have an unruly family!" said Mrs. Tabitha Twitchit.

"I'm not afraid of rats. I will help you to find him; and whip him too! What is all that soot in the fender?"

"The chimney wants sweeping—Oh, dear me, Cousin Ribby—now Moppet and Mittens are gone!"

"They have both got out of the cup-board!"

Treatment: Session 2

Ribby and Tabitha set to work to search the house thoroughly again. They poked under the beds with Ribby's umbrella, and they rummaged in cupboards. They even fetched a candle, and looked inside a clothes chest in one of the attics. They could not find anything, but once they heard a door bang and somebody scuttered downstairs.

"Yes, it is infested with rats," said Tabitha tearfully, "I caught seven young ones out of one hole in the back kitchen, and we had them for dinner last Saturday. And once I saw the old father rat—an enormous old rat, Cousin Ribby. I was just going to jump upon him, when he showed his yellow teeth at me and whisked down the hole."

"The rats get upon my nerves, Cousin Ribby," said Tabitha.

Ribby and Tabitha searched and searched. They both heard a curious roly-poly noise under the attic floor. But there was nothing to be seen.

They returned to the kitchen. "Here's one of your kittens at least," said Ribby, dragging Moppet out of the flour barrel.

They shook the flour off her and set her down on the kitchen floor. She seemed to be in a terrible fright.

"Oh! Mother, Mother," said Moppet, "there's been an old woman rat in the kitchen, and she's stolen some of the dough!"

The two cats ran to look at the dough pan. Sure enough there were marks of little scratching fingers, and a lump of dough was gone!

Treatment: Session 3

"Which way did she go, Moppet?"

But Moppet had been too much frightened to peep out of the barrel again.

Ribby and Tabitha took her with them to keep her safely in sight, while they went on with their search.

They went into the dairy.

The first thing they found was Mittens, hiding in an empty jar.

They tipped up the jar, and she scrambled out.

"Oh, Mother, Mother!" said Mittens—

"Oh! Mother, Mother, there has been an old man rat in the dairy—a dreadful enormous big rat, Mother; and he's stolen a pat of butter and the rolling-pin."

Ribby and Tabitha looked at one another.

"A rolling-pin and butter! Oh, my poor son Thomas!" exclaimed Tabitha, wringing her paws.

"A rolling-pin?" said Ribby. "Did we not hear a roly-poly noise in the attic when we were looking into that chest?"

Ribby and Tabitha rushed upstairs again. Sure enough the roly-poly noise was still going on quite distinctly under the attic floor.

"This is serious, Cousin Tabitha," said Ribby. "We must send for John Joiner at once, with a saw."

Now this is what had been happening to Tom Kitten, and it shows how very unwise it is to go up a chimney in a very old house, where a person does not know his way, and where there are enormous rats.

Tom Kitten did not want to be shut up in a cupboard. When he saw that his mother was going to bake, he determined to hide.

Treatment: Session 4

He looked about for a nice convenient place, and he fixed upon the chimney.

The fire had only just been lighted, and it was not hot; but there was a white choky smoke from the green sticks. Tom Kitten got upon the fender and looked up. It was a big old-fashioned fireplace.

The chimney itself was wide enough inside for a man to stand up and walk about. So there was plenty of room for a little Tom Cat.

He jumped right up into the fireplace, balancing himself upon the iron bar where the kettle hangs.

Tom Kitten took another big jump off the bar, and landed on a ledge high up inside the chimney, knocking down some soot into the fender.

Tom Kitten coughed and choked with the smoke; and he could hear the sticks beginning to crackle and burn in the fireplace down below. He made up his mind to climb right to the top, and get out on the slates, and try to catch sparrows.

"I cannot go back. If I slipped I might fall in the fire and singe my beautiful tail and my little blue jacket."

The chimney was a very big old-fashioned one. It was built in the days when people burnt logs of wood upon the hearth.

The chimney stack stood up above the roof like a little stone tower, and the daylight shone down from the top, under the slanting slates that kept out the rain.

Treatment: Session 5

Tom Kitten was getting very frightened! He climbed up, and up, and up.

Then he waded sideways through inches of soot. He was like a little sweep himself.

It was most confusing in the dark. One flue seemed to lead into another.

There was less smoke, but Tom Kitten felt quite lost.

He scrambled up and up; but before he reached the chimney top he came to a place where

somebody had loosened a stone in the wall. There were some mutton bones lying about—

"This seems funny," said Tom Kitten. "Who has been gnawing bones up here in the chimney? I wish I had never come! And what a funny smell! It is something like mouse; only dreadfully strong. It makes me sneeze," said Tom Kitten.

He squeezed through the hole in the wall, and dragged himself along a most uncomfortably tight passage where there was scarcely any light.

He groped his way carefully for several yards; he was at the back of the skirting-board in the attic, where there is a little mark ✳ in the picture.

All at once he fell head over heels in the dark, down a hole, and landed on a heap of very dirty rags.

When Tom Kitten picked himself up and looked about him—he found himself in a place that he had never seen before, although he had lived all his life in the house.

Treatment: Session 6

It was a very small stuffy fusty room, with boards, and rafters, and cobwebs, and lath and plaster.

Opposite to him—as far away as he could sit—was an enormous rat.

"What do you mean by tumbling into my bed all covered with smuts?" said the rat, chattering his teeth.

"Please sir, the chimney wants sweeping," said poor Tom Kitten.

"Anna Maria! Anna Maria!" squeaked the rat. There was a pattering noise and an old woman rat poked her head round a rafter.

All in a minute she rushed upon Tom Kitten, and before he knew what was happening—

His coat was pulled off, and he was rolled up in a bundle, and tied with string in very hard knots.

Anna Maria did the tying. The old rat watched her and took snuff. When she had finished, they both sat staring at him with their mouths open.

"Anna Maria," said the old man rat (whose name was Samuel Whiskers),—"Anna Maria, make me a kitten dumpling roly-poly pudding for my dinner."

"It requires dough and a pat of butter, and a rolling-pin," said Anna Maria, considering

Tom Kitten with her head on one side.

"No," said Samuel Whiskers, "make it properly, Anna Maria, with breadcrumbs."

"Nonsense! Butter and dough," replied Anna Maria.

The two rats consulted together for a few minutes and then went away.

Treatment: Session 7

Samuel Whiskers got through a hole in the wainscot, and went boldly down the front staircase to the dairy to get the butter. He did not meet anybody.

He made a second journey for the rolling-pin. He pushed it in front of him with his paws, like a brewer's man trundling a barrel.

He could hear Ribby and Tabitha talking, but they were busy lighting the candle to look into the chest.

They did not see him.

Anna Maria went down by way of the skirting-board and a window shutter to the kitchen to steal the dough.

She borrowed a small saucer, and scooped up the dough with her paws.

She did not observe Moppet.

While Tom Kitten was left alone under the floor of the attic, he wriggled about and tried to mew for help.

But his mouth was full of soot and cobwebs, and he was tied up in such very tight knots, he could not make anybody hear him.

Except a spider, which came out of a crack in the ceiling and examined the knots critically, from a safe distance.

It was a judge of knots because it had a habit of tying up unfortunate blue-bottles. It did not offer to assist him.

Tom Kitten wriggled and squirmed until he was quite exhausted.

Presently the rats came back and set to work to make him into a dumpling. First they smeared him with butter, and then they rolled him in the dough.

Treatment: Session 8

"Will not the string be very indigestible, Anna Maria?" inquired Samuel Whiskers.

Anna Maria said she thought that it was of no consequence; but she wished that Tom Kitten would hold his head still, as it disarranged the pastry. She laid hold of his ears.

Tom Kitten bit and spat, and mewed and wriggled; and the rolling-pin went roly-poly, roly; roly, poly, roly. The rats each held an end.

"His tail is sticking out! You did not fetch enough dough, Anna Maria."

"I fetched as much as I could carry," replied Anna Maria.

"I do not think"—said Samuel Whiskers, pausing to take a look at Tom Kitten — "I do NOT think it will be a good pudding. It smells sooty."

Anna Maria was about to argue the point, when all at once there began to be other sounds up above—the rasping noise of a saw; and the noise of a little dog, scratching and yelping!

The rats dropped the rolling-pin, and listened attentively.

"We are discovered and interrupted, Anna Maria; let us collect our property—and other people's, — and depart at once."

"I fear that we shall be obliged to leave this pudding."

"But I am persuaded that the knots would have proved indigestible, whatever you may urge to the contrary."

Testing: Post-test

"Come away at once and help me to tie up some mutton bones in a counterpane," said Anna Maria. "I have got half a smoked ham hidden in the chimney."

So it happened that by the time John Joiner had got the plank up—there was nobody under the floor except the rolling-pin and Tom Kitten in a very dirty dumpling!

But there was a strong smell of rats; and John Joiner spent the rest of the morning sniffing and whining, and wagging his tail, and going round and round with his head in the hole like a gimlet.

Then he nailed the plank down again, and put his tools in his bag, and came downstairs.

The cat family had quite recovered. They invited him to stay to dinner.

The dumpling had been peeled off Tom Kitten, and made separately into a bag pudding, with currants in it to hide the smuts.

They had been obliged to put Tom Kitten into a hot bath to get the butter off.

John Joiner smelt the pudding; but he regretted that he had not time to stay to dinner, because he had just finished making a wheel-barrow for Miss Potter, and she had ordered two hen-coops.

And when I was going to the post late in the afternoon—I looked up the lane from the corner, and I saw Mr. Samuel Whiskers and his wife on the run, with big bundles on a little wheel-barrow, which looked very much like mine.

Testing: Delayed post-test

They were just turning in at the gate to the barn of Farmer Potatoes.

Samuel Whiskers was puffing and out of breath. Anna Maria was still arguing in shrill tones.

She seemed to know her way, and she seemed to have a quantity of luggage.

I am sure *I* never gave her leave to borrow my wheel-barrow!

They went into the barn, and hauled their parcels with a bit of string to the top of the hay mow.

After that, there were no more rats for a long time at Tabitha Twitchit's.

As for Farmer Potatoes, he has been driven nearly distracted. There are rats, and rats, and rats in his barn! They eat up the chicken food, and steal the oats and bran, and make holes in the meal bags.

And they are all descended from Mr. and Mrs. Samuel Whiskers—children and grand-children and great great grand-children.

There is no end to them!

Moppet and Mittens have grown up into very good rat-catchers.

They go out rat-catching in the village, and they find plenty of employment. They charge

so much a dozen, and earn their living very comfortably.

They hang up the rats' tails in a row or the barn door, to show how many they have caught—dozens and dozens of them.

But Tom Kitten has always been afraid of a rat; he never durst face anything that is bigger than—

A Mouse.

Appendix D

Coding of Testing Passages

Pre-test Passage (M = Main idea; D = Detail / Non main idea)

Hierarchical level (M, D)	Weight (1, 2)	Number #	Idea unit
D	1	1	ONCE upon a time
M	2	2	there was an old cat,
M	2	3	called Mrs. Tabitha Twitchit,
M	2	4	who was an anxious parent.
M	2	5	She used to lose her kittens continually,
M	2	6	and whenever they were lost
M	2	7	they were always in mischief!
M	2	8	On baking day
M	2	9	she determined to shut them up
D	1	10	in a cupboard.
M	2	11	She caught Moppet and Mittens,
M	2	12	but she could not find Tom.
M	2	13	Mrs. Tabitha went up and down
M	2	14	all over the house,
D	1	15	mewing for Tom Kitten.
D	1	16	She looked in the pantry
D	1	17	under the staircase,
D	1	18	and she searched the best spare bedroom
D	1	19	that was all covered up
D	1	20	with dust sheets.
D	1	21	She went right upstairs
D	1	22	and looked into the attics,
M	2	23	but she could not find him anywhere.
M	2	24	It was an old, old house,
D	1	25	full of cupboards and passages.
D	1	26	Some of the walls were four feet thick,
D	1	27	and there used to be queer noises
D	1	28	inside them,
D	1	29	as if there might be a little secret staircase.
D	1	30	Certainly there were odd little jagged doorways
D	1	31	in the wainscot,
D	1	32	and things disappeared
D	1	33	at night—
D	1	34	especially cheese and bacon.
M	2	35	Mrs. Tabitha became more and more distracted,
D	1	36	and mewed dreadfully.

M	2	37	While their mother was searching the house,
M	2	38	Moppet and Mittens had got into mischief.
M	2	39	The cupboard door was not locked,
M	2	40	so they pushed it open
M	2	41	and came out.
M	2	42	They went straight to the dough
D	1	43	which was set to rise
D	1	44	in a pan
D	1	45	before the fire.
D	1	46	They patted it
D	1	47	with their little soft paws—
D	1	48	"Shall we make dear little muffins?"
D	1	49	said Mittens to Moppet.

Post-test Passage (M = Main idea; D = Detail / Non main idea)

Hierarchical level (M, D)	Weight (1, 2)	Number #	Idea unit
D	1	1	"Come away
D	1	2	at once
D	1	3	and help me
D	1	4	to tie up some mutton bones
D	1	5	in a counterpane,"
D	1	6	said Anna Maria.
D	1	7	"I have got half a smoked ham
D	1	8	hidden in the chimney."
D	1	9	So it happened
M	2	10	that by the time John Joiner had got the plank up—
M	2	11	there was nobody
M	2	12	under the floor
M	2	13	except the rolling-pin and Tom Kitten
D	1	14	in a very dirty dumpling!
M	2	15	But there was a strong smell of rats;
M	2	16	and John Joiner spent the rest of the morning
D	1	17	sniffing and whining,
D	1	18	and wagging his tail,
D	1	19	and going round and round
D	1	20	with his head
D	1	21	in the hole
D	1	22	like a gimlet.
M	2	23	Then he nailed the plank down again,
D	1	24	and put his tools
D	1	25	in his bag,
D	1	26	and came downstairs.

M	2	27	The cat family had quite recovered.
M	2	28	They invited him
M	2	29	to stay to dinner.
M	2	30	The dumpling had been peeled off Tom Kitten,
D	1	31	and made separately
D	1	32	into a bag pudding,
D	1	33	with currants in it
D	1	34	to hide the smuts.
D	1	35	They had been obliged
D	1	36	to put Tom Kitten
D	1	37	into a hot bath
D	1	38	to get the butter off.
D	1	39	John Joiner smelt the pudding;
D	1	40	but he regretted
M	2	41	that he had not time
M	2	42	to stay to dinner,
D	1	43	because he had just finished
D	1	44	making a wheel-barrow
D	1	45	for Miss Potter,
D	1	46	and she had ordered two hen-coops.
M	2	47	And when I was going
M	2	48	to the post
D	1	49	late in the afternoon—
M	2	50	I looked up the lane
D	1	51	from the corner,
M	2	52	and I saw Mr. Samuel Whiskers and his wife
M	2	53	on the run,
M	2	54	with big bundles
M	2	55	on a little wheel-barrow,
D	1	56	which looked very much like mine.

Delayed Post-test Passage (M = Main idea; D = Detail / Non main idea)

Hierarchical level (M, D)	Weight (1, 2)	Number #	Idea unit
M	2	1	They were just turning in
D	1	2	at the gate
M	2	3	to the barn
M	2	4	of Farmer Potatoes.
D	1	5	Samuel Whiskers was puffing
D	1	6	and out of breath.
D	1	7	Anna Maria was still arguing
D	1	8	in shrill tones.
D	1	9	She seemed to know her way,
D	1	10	and she seemed to have a quantity of luggage.

D	**1**	11	I am sure
M	**2**	12	I never gave her leave
M	**2**	13	to borrow my wheel-barrow!
M	**2**	14	They went into the barn,
M	**2**	15	and hauled their parcels
D	**1**	16	with a bit of string
D	**1**	17	to the top
D	**1**	18	of the hay mow.
M	**2**	19	After that,
M	**2**	20	there were no more rats
M	**2**	21	for a long time
M	**2**	22	at Tabitha Twitchit's.
M	**2**	23	As for Farmer Potatoes,
M	**2**	24	he has been driven nearly distracted.
M	**2**	25	There are rats, and rats, and rats
M	**2**	26	in his barn!
D	**1**	27	They eat up the chicken food,
D	**1**	28	and steal the oats and bran,
D	**1**	29	and make holes
D	**1**	30	in the meal bags.
M	**2**	31	And they are all descended
M	**2**	32	from Mr. and Mrs. Samuel Whiskers—
D	**1**	33	children and grand-children and great great grand-children.
D	**1**	34	There is no end to them!
M	**2**	35	Moppet and Mittens have grown up
M	**2**	36	into very good rat-catchers.
M	**2**	37	They go out rat-catching
D	**1**	38	in the village,
D	**1**	39	and they find plenty of employment.
D	**1**	40	They charge so much a dozen,
D	**1**	41	and earn their living very comfortably.
D	**1**	42	They hang up the rats' tails
D	**1**	43	in a row
D	**1**	44	on the barn door,
D	**1**	45	to show how many they have caught—
D	**1**	46	dozens and dozens of them.
M	**2**	47	But Tom Kitten has always been afraid of a rat;
M	**2**	48	he never durst face anything
M	**2**	49	that is bigger than—
M	**2**	50	A Mouse.