A Young Park*, Talia Isaacs and Helen Woodfield

A comparison of the effects of extensive and intensive reading approaches on the vocabulary development of Korean secondary EFL learners

https://doi.org/10.1515/applirev-2017-0025

Abstract: Although language experts have long advocated the use of Extensive Reading (ER) to enhance vocabulary acquisition, the widespread use of the more traditional Intensive Reading (IR) approach prevails in English as Foreign Language (EFL) settings. Many experimental studies have attempted to demonstrate the benefits of using ER over IR in classroom contexts; however, none have demonstrated significant differences in learning gains. This quasi-experimental study involved the measurement of partial vocabulary knowledge of specific words encountered through reading to compare the effects of ER and conventional IR instruction on EFL learners' vocabulary knowledge development. Two intact classes of 72 Korean secondary students (36/class) received either ER or IR instructional treatments over a 12-week timespan, with pre- and post- performance differences examined by proficiency level. ANCOVA results showed that students benefited significantly more from the ER than from the IR treatment in terms of their knowledge of the meanings and uses of target words. With regard to proficiency, advanced and intermediate level learners benefited more from ER, while low level learners benefited more from IR. These findings suggest that EFL practitioners should carefully consider their learners' proficiency level when selecting a reading approach, in order to optimize learners' vocabulary development.

Keywords: extensive reading, intensive reading, EFL learners, proficiency levels, vocabulary knowledge

1 Introduction

Numerous SLA scholars have championed Extensive Reading (ER) in promoting the vocabulary acquisition of second language (L2) learners for over three

***Corresponding author: A Young Park,** Changwon Girls' Highschool, Changwon, Korea, E-mail: sonnik@hanmail.net
Talia Isaacs, University College London, London, UK, E-mail: talia.isaacs@ucl.ac.uk
Helen Woodfield, Faculty of Social Sciences and Law, University of Bristol, Bristol, UK, E-mail: Helen.Woodfield@bristol.ac.uk

decades (Nakanishi 2015; Nation 2015; Day and Prentice 2016), arguing that this approach encourages learners to read large amounts of long, easy-to-understand material based on each individual learner's interest and reading level. Although a series of experimental studies have demonstrated a positive effect of ER on vocabulary acquisition (e.g. Cha 2009; Rashidi and Piran 2011; Eckerth and Tavakoli 2012; Song and Sardegna 2014), conventional Intensive Reading (IR) has continued to dominate L2 classroom settings, particularly in English as a Foreign Language (EFL) contexts, where input in the target language is limited (Renandya and Farrell 2011). The IR approach, which is teacher led, aims to support L2 learners in constructing detailed meaning from a reading text through close analysis and translation, in order to develop their linguistic knowledge (Nation 2013; Day and Prentice 2016).

Although there have been a few attempts to provide empirical evidence of the benefits of the ER approach on L2 vocabulary development (e.g. Rashidi and Piran 2011; Senoo and Yonemoto 2014; Song and Sardegna 2014), no comparative studies between ER and IR have, as yet, demonstrated a significantly positive effect of ER over IR. However, the trend towards significance observed in the above cited studies could suggest that methodological limitations of measuring vocabulary knowledge, as opposed to problems with the ER approach itself, may have accounted for the null result (Webb and Chang 2015). For example, the measurement tools used in most of these studies consisted of ready-made standardized vocabulary tests, such as the Vocabulary Size Test and the Vocabulary Level Test (e.g. Rashidi and Piran 2011; Elgort and Warren 2014; Webb and Chang 2015). However, these standardized tests measure learners' general vocabulary size rather than testing the words that they actually encountered through their reading during the ER intervention.

As Grabe and Stoller (2009) argue, teaching reading is "a complex matter" (p. 4) with individual differences in learner characteristics potentially affecting the degree of the success of L2 reading instructional interventions – for example, in their acquisition of lexis (Zahar et al. 2001; Stephens 2016). However, surprisingly few ER studies have been conducted that take language proficiency into consideration (e.g. Lee and Schallert 1997; Tekmen and Daloğlu 2006; Rashidi and Piran 2011; Park 2017). A rare exception is Rashidi and Piran's (2011) investigation of the effects of ER and IR on Iranian adult EFL learners' English vocabulary development as a function of their L2 English proficiency level, with significant differences detected between levels. However, this study covered a limited range of proficiency levels (advanced and high-advanced), which may not generalize to mixed classrooms with a wider proficiency range. Another exception is a study by Zahar et al. (2001), which investigated vocabulary gains in relation to a relatively wider range of proficiency levels (low,

intermediate and advanced). However, this study focused on L2 learners' incidental vocabulary learning through reading of a single reading text rather than the impact of the ER approach on vocabulary development across multiple reading texts chosen by learners. Therefore, there remains a need to investigate the instructional effectiveness of the ER versus the IR approach for learners from wide ranging proficiency levels, particularly in EFL classroom settings, where class sizes tend to be large, and where mixed-proficiency level classes are common (Alavinia and Farhady 2012).

In light of these gaps, the goal of the current study was to compare the impact of the ER and IR approaches on L2 vocabulary development by testing words actually contained within the reading texts. A further goal was to extend previous research on the impact of the ER and IR approaches by examining learners at a range of proficiency levels (advanced, intermediate, low) to investigate which learners stand to benefit the most from each instructional approach.

2 Background

2.1 Vocabulary learning

Research has shown that vocabulary knowledge is a key component of L2 competence (Daller and Phelan 2013; Roche and Harrington 2013; Daller and Yixin 2016). Consequently, the mechanisms through which L2 learners build vocabulary knowledge have been the subject of systematic research for over a century (Treffers-Daller and Milton 2013). In the context of first language (L1) acquisition, vocabulary knowledge is mainly acquired through natural, communicative exposure without much need for explicit vocabulary instruction (Senoo and Yonemoto 2014). Conversely, in L2 contexts, vocabulary learning tends to take place through explicit vocabulary instruction or incidental exposure to written input (Nation 2013; Reynolds 2015). Whereas explicit vocabulary instruction facilitates L2 lexical learning by efficiently highlighting form-meaning relationships (Elgort and Warren 2014), incidental vocabulary acquisition that takes place through reading has at least two benefits. Firstly, vocabulary acquisition and the act of reading occur simultaneously (Nation 2015), which is more efficient from a pedagogical perspective. Secondly, in incidental vocabulary acquisition, a richer sense of a word can be gained through diverse contextualized input (Nation 2013). This is because whenever a word occurs in a reading text, it typically occurs in a different context each time. Each new encounter with a word during incidental vocabulary acquisition is likely to enrich overall

knowledge of that word through its varied contexts (Nation 2015; Day and Prentice 2016). Thus, incidental vocabulary acquisition can offer learners the opportunity to encounter words in their contexts of use. Accordingly, contextual appropriateness of words can be acquired from ER in addition to word meanings. These advantages strengthen the position taken by numerous L2 reading researchers in recommending the ER approach as the preferred instructional technique to enhance L2 learners' vocabulary knowledge (Webb and Macalister 2013; Song and Sardegna 2014; Nation 2015). By contrast, the IR approach entails the intentional and conscious learning of form-meaning relationships of target words, making it time-efficient in increasing learners' vocabulary size (Eckerth and Tavakoli 2012). This is one of the main reasons why the IR approach has traditionally monopolized EFL classrooms in contexts where limited English input is available (Negi 2010).

2.2 Vocabulary tests in ER studies

As mentioned earlier, while a few studies have shown that ER interventions are as effective as IR in improving participants' vocabulary test performance, none have, as yet, yielded a statistically significant positive effect for ER relative to IR. However, this lack of significant improvement may have resulted from measurement issues to do with operationalizing constructs rather than from problem with ER itself (Horst 2005). For example previous ER studies likely use ready-made standardized vocabulary measures to measure pre- and post-intervention vocabulary due to logistical challenges. Specifically, because learners self-select their own reading texts in ER lessons, there is no single text that all learners have read as a collective and, therefore, no one set of words that they can uniformly be tested on before and after an ER intervention (Park 2017). However, a number of studies on incidental vocabulary learning through ER have made use of a single reading text rather than multiple reading texts that participants choose to read (e.g. Zahar et al. 2001; Pigada and Schmitt 2006; Vidal 2011). As Webb and Chang (2015) argue, this is problematic because these studies have not captured the essence of incidental vocabulary learning through ER. The reason for their claim is that "encountering unknown words n times in a single text may lead to a different degree of vocabulary learning than encountering unknown words the same number of times in multiple texts" (Webb and Chang 2015: 668). There have been rare attempts within ER research to develop vocabulary measures consisting of word items from multiple self-selected reading texts, and these attempts have found significant gains in vocabulary knowledge from ER (e.g. Horst 2005; Kweon and Kim 2008).

However, the scope of these studies was limited to the impact of ER alone, without comparing it to IR.

In addition, in previous ER studies, partial vocabulary knowledge, which refers to "different levels of knowledge of words" (Read 1993: 357), has been overlooked in terms of the lexical performance measures that were used (e.g., Cha 2009; Webb and Chang 2015). Capturing partial vocabulary knowledge is especially important in ER research due to the repeated opportunities to encounter and, thereby, gain varying degrees of knowledge of previously unknown lexical items. This incidentally acquired aspect of vocabulary knowledge is likely to be strengthened through ER. Thus, not using partial vocabulary knowledge measures to examine learner gains is a methodological shortcoming. In fact, few existing ER studies have moved beyond assessing vocabulary knowledge through word meaning, despite the fact that ER has long been recognized as contributing to multiple aspects of vocabulary knowledge such as word usage (Song and Sardegna 2014). This is a matter for consideration because, as Webb (2005) argues, "studies that measure only word meaning may be unable to find a significant result when there have in fact been significant gains in other aspects of vocabulary knowledge" (p. 48).

To summarize, further research is needed to compare the impact of the ER and IR approaches on vocabulary knowledge, particularly as the failings by previous studies to detect a significant advantage for ER may have been due to imprecise measurement of learners' attainment (Webb and Chang 2015). Therefore, to more accurately quantify the benefits of the ER approach on partial vocabulary knowledge in terms of meaning and use, a more sensitive vocabulary test was developed in the current study that directly tested lexical items encountered through multiple reading texts from the ER approach. The next section will probe the issue of learner proficiency more deeply, as different pedagogical approaches may benefit learners more or less in relation to their proficiency level.

2.3 Language proficiency

In order to outline the relationship between L2 learners' proficiency and their linguistic development in the current study, it is necessary to briefly review the linguistic threshold hypothesis. This hypothesis holds that L2 learners need to obtain a threshold level of L2 proficiency to achieve functional L2 competence (Hudson 2007; Grabe 2009), thereby underscoring a central role for L2 proficiency in L2 reading pedagogy (Stephens 2016). According to Lee and Schallert (1997), who extend the linguistic threshold hypothesis to the L2 reading context, L2 learners must first reach a threshold level of L2 proficiency in order to become

fluent L2 readers (Grabe and Stoller 2009; Laufer and Ravenhorst-Kalovski 2010). Grabe and Stoller (2009) propose that L2 learners cross this threshold "when they encounter L2 text in which they know almost all of the words and can process the text fluently" (p. 51). It is now widely accepted that the minimum lexical coverage needed for L2 reading comprehension is the point at which the reader meets the knowledge of 4000–5000 word families yielding the coverage of 95% known-word criterion (Schmitt et al. 2011).

The linguistic threshold hypothesis is in line with the "beginner's paradox" postulated by Coady (1997). According to "Beginner's paradox", "there is a threshold level of vocabulary knowledge below which a learner cannot read well enough to learn new vocabulary through reading" (Coady 1997: 229). Conversely, an L2 learner with a proficiency level above the threshold may be able to extend his or her vocabulary through incidental vocabulary acquisition by reading extensively. In short, the linguistic threshold hypothesis highlights that L2 reading and vocabulary development are closely related to L2 proficiency (Milton 2013; Daller and Yixin 2016). However, to date, little research has directly compared the ER and IR approaches in relation to L2 learners' differential proficiency in classroom settings (cf, Rashidi and Piran 2011; Yamashita 2015; Park 2017).

This research gap suggests the need for further ER research in relation to learners' proficiency, particularly in EFL contexts. Despite being a group that could benefit from the ER approach, EFL learners have been underresearched compared to learners in second language contexts (Park 2017). Since EFL teachers often struggle to fulfill the different learning needs of their students in mixed-proficiency classes, they often try to 'pitch' the class to learners' average proficiency level (Negi 2010). However, this uniform approach to teaching students at different levels could lower learners' motivation (Nation 2013). For example, in mixed-level classes, it is common that "advanced-proficiency level students feel bored and low-proficiency level students feel discouraged and overwhelmed" (Negi 2010: 89). Nonetheless, due to the challenges inherent in teaching large classes that encompass a wide range of L2 proficiency levels, resorting to using uniform teaching approaches that neglect diverse proficiency levels is often the default option for teachers (Park 2017). This uniform teaching approach has also been applied to reading lessons in EFL secondary school classes by using a one-size-fits-all reading approach (i.e. IR) that fails to take diverse proficiency levels into account (Glew 2013). To sum up, because L2 proficiency is such a critical factor in L2 reading pedagogy, further investigation into learners' proficiency level in relation to L2 reading is needed, particularly in underresearched EFL contexts that need to deal with large class sizes and where the IR approach persists (Day and Prentice 2016). Therefore, the research questions (RQs) for the present study are as follows:

- RQ1. To what extent do ER and IR approaches effect Korean secondary EFL learners' vocabulary development over a 12-week instructional period?
- RQ2. Do Korean secondary learners from different L2 English proficiency levels differentially benefit from IR versus ER approaches?

3 Methods

3.1 Participants

This quasi-experimental study included two all-female suburban private secondary school classes ($M_{Age} = 15.3$ years; range = 15–16). All participants had 6 to 10 years of experience of learning English officially or informally. Their English proficiency was diverse from low to advanced level. Their TOEIC scores were an average of 467.2, which is an intermediate proficiency level. One class of 36 students received ER lessons (experimental group 1) and another class of 37 students, matched for overall proficiency level, received traditional IR lessons (experimental group 2). However, one participant from the IR group was excluded, as she had lived in an English-speaking country for two years. As a result, only 36 out of the 37 participants' data were used for the IR group, which matched the 36 participants in the ER group. Both lessons were taken once a week for two hours for one academic semester lasting 12 weeks and were taught by the same teacher. Using a mock College Scholastic Ability Test (CSAT[1]) for English, a placement test administered prior to the beginning of the school semester and before the start of the study. Students in both groups had been divided into three subgroups according to their L2 English proficiency level: low, intermediate and advanced, and equally interspersed into the two classes. The researchers had access to these placement test data and, as part of the current study, used their proficiency level classification to group them into beginner, intermediate and advanced levels.

3.2 Treatment

The experimental (ER) group received lessons based on a class library (a full set of 50 graded readers) for students to borrow. Graded readers are written with

1 The mock College Scholastic Ability Test (CSAT) is administered four times a year by 16 Metropolitan and Provincial Offices of Education in Korea. It adopts the same format and difficulty level as the CSAT, which is the high-stakes university entrance test approved by the Korean Ministry of Education (Shin et al. 2011).

carefully controlled vocabulary and sentence structures to provide progressive difficulty and complexity for different levels of readers (Day and Prentice 2016). Participants were advised to read one graded reader per week of their own choice for 12 weeks, conforming to guidelines about the number of repetitions needed to reinforce the meaning of new words from the vocabulary literature (Nation 2015). Each devised an individualized reading list of books from the graded readers collection that they wished to read over the 12-week ER course, totaling 12 graded readers per participant. To promote incidental vocabulary learning, they were instructed to skip unknown words if doing so would not interfere with their understanding of the story (Waring and Takaki 2003). Unlike IR lessons, no instruction relating to reading skills and strategies was targeted in their regular EFL instruction. After each lesson, participants in the ER group were instructed to complete the rest of their reading as homework. They were also required to keep a reading log to monitor their reading. According to their reading logs, participants spent an average time of 164 minutes ($SD = 91$) reading outside of class each week.

The IR lessons consisted of carefully structured and phased instruction involving analyzing and translating a reading text once a week for two hours. During each lesson, the IR group studied four short reading texts from the chosen reading material consisting of approximately 700 to 800 words. The reading material used was a reading exercise book called "Finalizing English Reading I" (Kim 2010). This book had been selected for the upcoming term by the school's English teachers when setting the curriculum, in accordance with the list of Ministry of Education guidelines in Korea, before term started. The IR lessons involved a reading exercise book consisting of short reading texts at the participants' "instructional reading level", that is, "where they need some help to understand the reading material" (Jacobs and Farrell 2012: 2). Therefore, close guidance from the teacher was required for the IR group to understand the detailed meaning of the reading text (Rashidi and Piran 2011). To guide participants' reading, the teacher offered explicit instruction of reading strategies, grammar points and vocabulary relating to the texts. During the lesson, the teacher instructed the IR group to read four texts from the reading exercise book and helped them complete the accompanying exercises. As an assignment, they were required to read four new reading texts from the exercise book per week and to complete further exercises simulating what these participants would have been familiar with in their regular IR lessons. For their reading assignments, IR group read an average of 192 minutes outside of class during the treatment. To confirm that the participants had completed the reading assignments, a series of check-up quizzes was administered at the beginning of the following class.

3.3 Data collection instruments

3.3.1 Developed vocabulary test

The current study developed a vocabulary test designed to capture the acquired vocabulary knowledge from the ER approach in a more sensitive way than the measures employed in previous studies. This was an attempt to address previous studies' methodological limitations in regard to (i) testing participants on vocabulary items that they may not have encountered while reading and (ii) measuring general vocabulary size rather than partial vocabulary knowledge in terms of meaning and use. Thus, the current study sought to develop a test closely modeled on Horst's (2005) vocabulary test to capture partial vocabulary knowledge of specific words encountered by learners in their reading. However, a limitation of Horst's (2005) test was that it only partially represented the participants' exposure to new vocabulary by including only vocabulary items from the first 20 pages from each book. In addition, Horst's (2005) vocabulary test adopted the Vocabulary Knowledge Scale (VKS; Paribakht and Wesche 1997), a limitation of which is that mean VKS scores for each participating group do not explicitly illustrate how learners perform on individual target words (Bruton 2009). To elaborate, participants' individual scores are ambiguous in that two similar scores might be the result of completely different combinations of responses. For example, not knowing four words at all (equivalent to a total score of 4 points using the VKS) is equal in value to knowing one word to the level of being able to provide the correct definition (also equivalent to a total score of 4 points). The format of the VKS was deemed suitable for capturing learners' vocabulary development in this study, since the VKS nonetheless reflects an understanding of vocabulary acquisition as being an iterative process that involves different levels and aspects of knowledge (Read 2000). However, the point cited above that participants' total scores do not specify which scoring combinations provided the overall result must also be considered (Bruton 2009).

Accordingly, the current study attempted to overcome the limitations of Horst's (2005) vocabulary test by incorporating the following design features. First, the vocabulary test developed here fully represented all the vocabulary items encountered by the participants. This was achieved by electronically scanning the entire text of every graded reader used in the intervention to design an individualized test tailored to each participant's choice of books. Thus, the developed test could fully reflect all word learning opportunities available to participants who had read the entire text of each chosen book. Secondly, the current study applied a modified scoring system of VKS which starts with 0

point for Category I, 1 point to the Category II, 2 points to the Category III and so on. That is, the modified VKS score system did not assign any point to word knowledge of Category I ("I don't remember having seen this word before") while original VKS score system assigned 1 point. The current study adopted a modified scoring system of VKS because this scoring system prevents participants' test scores from representing the sum of ambiguous value of target words and the sum of different criteria. Thirdly, in accordance with Kim (2008), the current study calculated the total number of words receiving each score on the VKS to clarify where changes in participants' vocabulary growth had taken place. Detailed descriptions of the test development process and the test structure related to the VKS format appear in the next section.

3.3.2 Instrument development process

3.3.2.1 Development of a sample test

Prior to developing the vocabulary test for the current study, a sample test for a pilot study was carried out in which 12 graded readers were randomly selected from diverse reading levels. This was done to simulate an individual participant's choice of 12 graded readers according to personal interest and proficiency level, as was planned for members of the ER group in the main study. Second, the 12 randomly selected graded readers were electronically scanned to create lists of the words that had appeared in each book. This pool of word lists was then used to create a different test for each participant based on her reading. Third, to strengthen their internal consistency, the words chosen for each participant's test were carefully examined to ensure that they contained comparable numbers of "difficult" (low frequency) and "easy" (high frequency) words. To achieve this, the software "Lexical Frequency Profiling" (LFP; Nation and Heatley 1996), which incorporates West's (1953) General Service List (GSL) and Coxhead's (2000) Academic Word List (AWL), was used to categorize the test words. This software divides the words into four frequency categories: (a) the 1000 most frequent word families of English (GSL 1000); (b) the 1001–2000 most frequent word families of English (GSL 2000); (c) 570 words families that frequently occur in university texts across academic disciplines (AWL); and (d) less frequent words and proper nouns that do not appear in any of the preceding categories but that were featured in participants' reading materials (off-list words). All words in the test pool were categorized using the "VocabProfile V1.5" function using Cobb's (2004) "Complete Lexical Tutor", the on-line version of the LFP software.

3.3.2.2 Pilot study

The sample test was piloted twice on five randomly selected Korean secondary school students drawn from the same population as in the main study in terms of English proficiency level (MIELTS scores = 5.3; range: 0–9). One hundred twenty lexical items were selected for each participant's choice of 12 books using an on-line version of List Randomizer, which automatically draws its randomizations from categories (Haahr and Haahr 2010). All words selected for each participant were categorized according to their frequency so as to be equivalent in difficulty level to the test words selected for all other participants. The first pilot revealed that the test was too easy for the participants, as the mean score achieved was 87.52/100 (range: 78–96). To avoid a ceiling effect, the difficulty level was increased through the addition of a greater proportion of difficult words (operationalized as low-frequency words) and concomitant reduction of easy (high-frequency) words (Millett et al. 2007). In addition, it took too long for the pilot participants to complete the test (M = 97 minutes; range: 81–105) in light of the limited class time available and to avoid fatigue effects. Therefore, the number of testing items was reduced from 120 to 70. In the subsequent pilot using the modified vocabulary test, the participants' mean test score was 72 out of 100 (range: 52–85) and it took them an average of 55 minutes (range: 50–78) to complete. In post-pilot interviews, the participants gave generally positive feedback on the adjusted test, commenting that the number of test items and difficulty level were appropriate. Based on these pilots, the finalized word list for the main study was set at 10 words from the GSL 1000, 25 from the GSL 2000, 15 from the AWL, and 20 off-list words. The words from the off-list category were reviewed to exclude typographical errors from the developed test. Finally, for the main data collection for ER group, a full set of 50 graded readers was scanned and 70 words were randomly selected from the word lists of all 12 books that each participant had chosen to read. Compared to the involved procedure detailed above, it was relatively easy to determine the target words for testing the IR group. Since all IR participants had read one identical provided text, target words were simply selected from it. That is, while the ER participants all read different books and then took customized tests derived from their individual reading books, all IR participants read the same book and then took the same test. The method for selecting words and categorizing their frequency was the same as for the ER group. Since the number of test items was quite high (70 items) and there were only a limited number of words to select from, the same test items were used for the pre- and post-tests in both groups, albeit presented in a different order.

3.3.3 Vocabulary test structure

The scoring system for the developed test adopted modified version of Paribakht and Wesche's (1997) VKS. In a test using the modified version of the VKS, participants were presented with a list of target words and were asked to indicate their level of knowledge of each word on a 5-level rating scale developed for the tool. The scoring system is: 0 point = "Category I: I haven't seen this word and I don't know what it means" (total unfamiliarity); 1 point = "Category II: I have seen this word before, but I don't know what it means"; 2 points = "Category III: I have seen this word before, and I think it means _____"; 3 points = "Category VI: I know this word. It means _____" (recognition of the word and some idea of its meaning); and finally, 4 points = " Category V: I can use this word in a sentence ____" (competence to use the word with grammatical and semantic accuracy in a sentence; Paribakht and Wesche 1997: 181). This scoring system was used in the present study to record participants' responses both before and after receiving the reading interventions.

In summary, to design an individualized test tailored to each participant's choice of books in the ER condition, electronically scanning the text of each graded reader enabled ER participants to be tested only on words actually encountered in their reading. This arguably enhanced the content validity of the developed vocabulary test relative to previous studies examining ER, which tested vocabulary that had no relation to the reading texts. In addition, the developed vocabulary test measured partial vocabulary knowledge; not only the meaning of the target word, but also its use in a sentence.

3.4 Procedure

Following completion of a background questionnaire, the developed vocabulary tests for the 12-week ER and IR treatments were used to collect pre- and post-test data. Participants were not informed that they would be taking the test again at the end of the experiment. Participants' responses were scored by two markers using a 5-point scoring guide designed by the first author. The markers were full-time secondary school English teachers with extensive English reading and writing proficiency test scoring experience in Korea.

3.5 Data analysis

For RQ1, a between-subjects t-test was first conducted to compare the mean pre-test scores of the ER and IR groups. Next, a one-way between-groups analysis of covariance (ANCOVA) was performed to control for initial group differences on

the pre-test. For RQ2, a two-way ANCOVA was conducted to investigate whether the treatment's impact on participants differed according to their proficiency level. In this analysis, there was one between-subjects independent variable with two levels, specifically, the type of treatment (ER or IR approach), and one within-subjects independent variable with three levels relating to English proficiency level (advanced, intermediate or low). The dependent variable was the post-test scores on the vocabulary test and the covariate was the pre-test scores. The probability value of $p \leq 0.05$ was adopted for all statistical tests.

4 Results

Before the t-test and ANCOVA were applied, preliminary checks were conducted. The two sets of ratings were averaged and the ratings made by judges showed acceptable inter-rater reliability (Cronbach's alpha = 0.75). Research question 1 compared the ER and IR reading approaches, investigating the effect of each on Korean secondary EFL learners' vocabulary development following a 12-week intervention. The t-test showed that the ER group's mean vocabulary pre-test score ($M = 195.07$, $SD = 13.38$) was higher than that of the IR group ($M = 189.1$, $SD = 20.51$), although statistical significance was not achieved $t\ (70) = 0.149$, $p = 1.46$, two-tailed. Nonetheless, to control for this initial between-group difference in terms of mean scores and statistically establish a common baseline between the classes at the outset of instruction, a one-way between-subjects ANCOVA was computed to examine whether there were differences between the groups on the post-test measure, with scores on the vocabulary pre-test applied as the covariate 192.08. ($p < 0.05$). ANCOVA results estimated that the ER group ($M = 212.28$) significantly outperformed the IR group ($M = 204.74$) in the vocabulary post-test, $f\ (1,\ 69) = 5.95$, $p = 0.02$, although the effect size was small ($\eta_p^2 = 0.08$). The ER group's post-test adjusted mean score of vocabulary knowledge increased significantly more than of that of the IR group by 7.54 ($p < 0.05$), as is shown in Figure 1.

To clarify how the learners performed on the individual test items, Table 1 shows each group's VKS scores broken down by the five VKS categories. It is worth noting the number of words that received a score of 3 or above on the VKS, since those scores indicated that learners knew not only meaning, but also were able to use the target word in context based on their self-assessment. Whereas the mean number of words rated 3 or above increased from 1686 to 1877 for the ER group, that of IR group increased from 1659 to 1672. That is, the

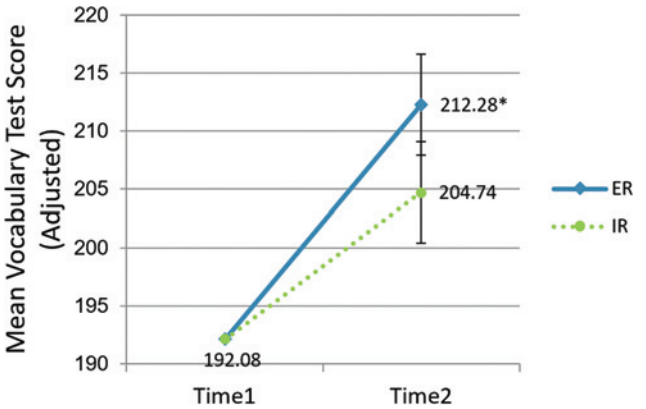

Figure 1: Estimated means on the vocabulary knowledge test at Time 2 (Post-Test) for the ER and IR groups.
Note: Covariate values of the vocabulary knowledge mean score at Time 1 (pre-test) in both groups was evaluated at 192.08. Brackets enclose ±1 SE. An asterisk designates a significant difference between the two groups using ANCOVA adjusted means ($p < 0.05$). The maximum possible score on the vocabulary knowledge test was 280 and the minimum was 0.

ER group's gains in the number of words which they knew in terms of meaning and use after the intervention was over 14 times higher than the collective gains made the IR group.

Taken together, the results show that when differential pre-test between-group performance on the vocabulary knowledge measure was controlled for using an ANCOVA, the ER group showed greater learning gains of vocabulary knowledge than did the IR group, in terms of meanings and uses of target words.

To address RQ2, a two-way ANCOVA was calculated to compare the impact of the ER and IR approaches on the vocabulary knowledge of secondary Korean EFL learners according to their English proficiency level, as assessed using the proficiency classification from mock CSAT. The independent variables were, therefore, treatment type (ER or IR approach) and proficiency level (advanced, intermediate or low) and the dependent variable was the vocabulary test mean score at Time 2 (post-test). Mean scores from the vocabulary test at Time 1 (pre-test) were used as the covariate to control for initial group differences. The ANCOVA revealed a statistically significant interaction effect between ER and IR approach, $f\ (2,\ 65) = 13.17$, $p < 0.05$; $\eta_p^2 = 0.29$. There was also a statistically significant main effect for proficiency level, $f\ (2,\ 65) = 12.58$, $p < 0.05$, $\eta_p^2 = 0.28$. That is, the three proficiency groups

benefited differently from the ER and the IR approach. The ER approach made a significantly more positive impact on advanced and intermediate proficiency groups' vocabulary knowledge compared to the IR approach, as shown in Figure 2.

5 Discussion

This classroom-based intervention study compared the effect of the ER and IR reading approaches on Korean secondary EFL learners' vocabulary acquisition, as measured through pre- and post- test scores on a vocabulary knowledge test targeting lexical items that they had encountered during their readings. In addition, the study examined L2 proficiency effects. In response to RQ1, which compared the effect of IR and ER approaches on short-term vocabulary development, results revealed a significant improvement in the ER compared to the IR group's vocabulary performance. This contrasts with findings from previous ER studies that either failed to establish the superiority of ER over IR in promoting learning gains (e.g. Hulstijn 1992; Waring and Takaki 2003; Al-Homoud and

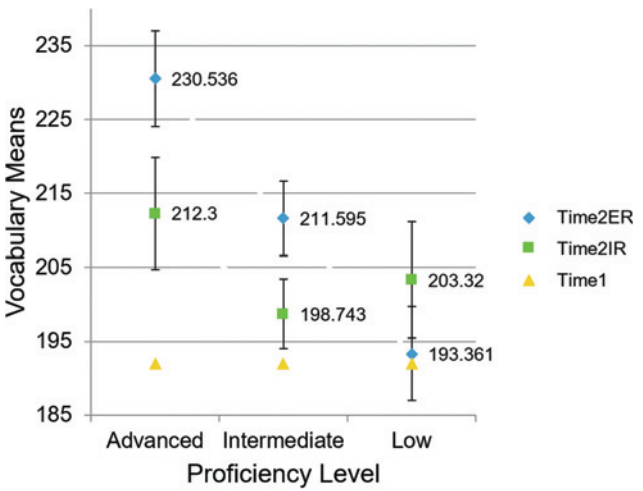

Figure 2: Estimated means on the vocabulary knowledge test for the ER and IR groups at Time 2 (Post-Test) using a two-way ANCOVA.
Note: Covariate values of the vocabulary knowledge mean score at Time 1 (pre-test) in both groups was evaluated at 192.08. Brackets enclose ± 1 SE. An asterisk designates a significant difference ($p < 0.05$). The maximum possible score on the vocabulary knowledge test was 280 and the minimum was 0.

Schmitt 2009). This lends credence to Horst's (2005) and Webb and Chang's (2015) assertions that the null result was likely due to shortcomings of the lexical measures used, which were not sensitive to content that the ER learners had actually encountered in their readings, rather than to the ER treatment itself. Thus, further research should seek to optimize the measures of vocabulary development that they examine along the lines of the custom-made lists developed in the present study, perhaps administering a combination of vocabulary measures and ideally with a delayed post-test to investigate whether learning gains are maintained over time.

Although the current study suggested that the ER approach was more conducive to promoting vocabulary learning compared to the IR approach through use of the developed test, it is unclear how learners performed on the individual test items due to the ambiguity of the VKS scoring system mentioned above. Therefore, the total number of words that received each score on the VKS was calculated. As Table 1 shows, the improvement in knowledge of words rated 4 or above in VKS categories suggests gains for the ER compared to the IR group in terms of word meaning and use. One possible explanation is that ER provides learners with a richer sense of a word's meaning and use through contextualized input than is provided by IR (Nation 2015).

With respect to the second research question, which asked how the ER and IR approaches differentially affect participants at different proficiency levels, results suggest that higher proficiency learners tend to benefit more from the ER approach, whereas there does not appear to be any advantage of ER over IR for lower level learners. More specifically, the results suggested that the advanced proficiency group benefited most from the ER treatment, with the intermediate group coming a close second. The low group gained the least from the ER approach, benefiting even less than the low IR group. Conversely, the low-

Table 1: Number of vocabulary items for each VKS score by reading intervention group (ER, IR).

VKS score	Time 1		Time 2	
	ER (*n*=36)	IR (*n*=36)	ER (*n*=36)	IR (*n*=36)
0	158(6.3 %)	130(5.2 %)	23(0.9 %)	86(3.4 %)
1	76(3.1 %)	419(16.6 %)	19(0.8 %)	59(2.3 %)
2	600(23.8 %)	312(12.4 %)	601(23.8 %)	703(27.9 %)
3	996(39.5 %)	864(34.3 %)	997(39.6 %)	867(34.4 %)
4	690(27.3 %)	795(31.5 %)	880(34.9 %)	805(32 %)
Total number of test words	2520	2520	2520	2520

Note: N = 36 participants in each group × 70 test items.

level IR group experienced the greatest benefit from the IR treatment, although they did not significantly outperform their low-level ER counterparts. This suggests that the linguist threshold theory is indeed applicable to the vocabulary knowledge measure operationalized in the present study. It is thus worth considering why, unlike in the ER group, the linguistic threshold theory did not apply to the findings from the IR group, which appear to run contrary to the linguistic threshold theory's predictions, with the lowest group improving most among the three proficiency levels. One possible explanation for this is the different emphasis on learner autonomy in ER and IR approaches (Yamashita 2015), with ER fostering a higher degree of autonomy by encouraging learners to read on their own without much help or guidance (Day and Prentice 2016). This learner-centered approach could be challenging or even intimidating or isolating for low proficiency learners. In contrast, in the teacher-centered IR approach, the teacher guides and helps the learners read and understand. Understandably, remaining under the close guidance of a teacher could allow low proficiency learners to feel more supported and confident, although this has yet to be empirically established.

6 Conclusion

The current study demonstrated that learners exposed to the ER intervention outperformed the IR group on measures of lexical knowledge and use. This finding stands out among existing ER studies' results, where no significant differences between the ER and IR approaches had previously been found. This highlights both the importance of using measures that align with the lexical exposure from reading materials, and the need to move beyond solely testing knowledge of word meaning to examining other dimensions of lexical knowledge and use. The results also show that the advanced group benefited most from the ER approach, whereas the low group benefited most from the IR approach in vocabulary development through the course of the instructional intervention. This can be at least partially explained by the linguistic threshold theory and degree of learner autonomy embedded in the ER approach.

The current study, which used intact classes matched for proficiency level before regular instruction, is grounded in the reality of classroom teaching and learning and the pervasiveness of IR over ER in pedagogical practice in many instructional contexts (Nation 2015). There are several implications for pedagogical practice. First, the current study revealed a positive effect of the ER approach compared to the IR approach on the vocabulary knowledge of participants at advanced and intermediate proficiency levels. This implies that the incidental vocabulary acquisition offered by the ER approach was more effective

for these learners, while the explicit vocabulary instruction offered by the IR approach was more beneficial for lower proficiency learners, lending support to the linguistic threshold theory (Grabe 2009; Park 2017). Therefore, when planning to teach vocabulary through reading, practitioners should consider whether the group's or, ideally, the individual student's L2 proficiency level lends itself more to the ER or IR approach. Moreover, practitioners should encourage low proficiency learners to focus their vocabulary learning on the most frequently used words, since this will help them to acquire the threshold level of vocabulary for reading comprehension before progressing to free reading interventions such as ER (Zhang and Anual 2008; Nation 2015).

A second related point is that learners with different proficiency levels benefit from different reading approaches. This suggests that EFL practitioners should consider streaming learners according to their English proficiency level in schools where student proficiency tends to be highly variable (Alavinia and Farhady 2012). Proficiency streaming allows practitioners to apply the reading approach that most effectively meets the learning needs and language level of the students in that particular class, which can enhance their reading fluency as well as vocabulary knowledge (Ireson and Hallam 2001).

Thirdly, although the current study demonstrated positive outcomes of the ER approach in an EFL secondary context, the successful implementation of ER in a real-world classroom setting poses practical challenges. One reason is because the student-centered ER approach requires teachers to adopt a new role with which they may be unfamiliar. That is, unlike within the teacher-centered IR approach, the role of teachers within the ER approach does not focus on delivering knowledge, but rather on supporting and encouraging learners to read independently (Day and Prentice 2016). Therefore, an in-service training course that takes the form of a mini ER course would be valuable, both in allowing practitioners to experience the benefits and logistical complexities of ER for themselves, and in helping them successfully apply the ER approach in their reading lessons.

Nevertheless, the limitations of this study, including the small sample size, short duration of treatment, lack of a control group, and lack of a post delayed-test, and differentiated reading time between the two groups, render this claim tentative rather than conclusive. In particular, it is important to note the limitation related to reading time difference. In the current study, the ER group spent a greater overall time reading than the IR group. This is because the nature of the ER approach required the ER group to read longer reading texts than the IR group. Consequently, it is possible that the ER groups' vocabulary knowledge gains were due not only to the difference in approach, but also to the additional reading time. While, from an experimental perspective, it would have been ideal

to control the time spent on task, from a pedagogical perspective it can be argued that the increased reading time necessitated by the ER approach is in itself a benefit, as Huffman (2014) underscores. Further research is needed to measure the long-term effect of the ER and IR approaches on more diverse aspects of vocabulary knowledge such as collocations, multiple word meanings (polysemy), and spellings of target words.

References

Alavinia, P. & S. Farhady. 2012. Using differentiated instruction to teach vocabulary in mixed ability classes with a focus on multiple intelligences and learning styles. *International Journal of Applied Science and Technology* 2(4). 72–82.

Al-Homoud, F. & N. Schmitt. 2009. Extensive reading in a challenging environment: A comparison of extensive and intensive reading approaches in Saudi Arabia. *Language Teaching Research* 13(4). 383–401. 10.1177/1362168809341508 (accessed 2 January 2017).

Bruton, A. 2009. The vocabulary knowledge scale: A critical analysis. *Language Assessment Quarterly* 6(4). 288–297. 10.1080/15434300902801909 (accessed 2 January 2017).

Cha, J. E. 2009. The effects of extensive reading on enhancing vocational high school students' L2 vocabulary and reading rates. *English Teaching* 64(3). 3–30. 10.15858/engtea.64.3.200909.3 (accessed 2 January 2017).

Coady, J. 1997. L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition: A rationale for pedagogy*, 225–237. Cambridge: Cambridge University Press.

Cobb, T. (2004). *The compleat lexical tutor*. Retrieved from http://www.lextutor.ca/

Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34(2). 213–238. 10.2307/3587951 (accessed 2 January 2017).

Daller, M. & D. Phelan. 2013. Predicting international student study success. *Applied Linguistics Review* 4(1). 173–193.

Daller, M. & W. Yixin. 2016. Predicting study success of international students. *Applied Linguistics Review* https://doi.org (accessed 2 January 2017).

Day, R. & N. Prentice. 2016. *Extensive reading into the classroom*. Oxford: Oxford University Press.

Eckerth, J. & P. Tavakoli. 2012. The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research* 16(2). 227–252. 10.1177/1362168811431377 (accessed 2 January 2017).

Elgort, I. & P. Warren. 2014. L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning* 64(2). 365–414. 10.1111/lang.12052 (accessed 2 January 2017).

Glew, P. J. 2013. Embedding international benchmarks of proficiency in English in undergraduate nursing programs: Challenges and strategies in equipping culturally and linguistically diverse students with English as an additional language for nursing in Australia. *Collegian* 20(2). 101–108. 10.1016/j.colegn.2012.04.002 (accessed 2 January 2017).

Grabe, W. 2009. *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
Grabe, W. & F. L. Stoller. 2009. *Teaching and researching: Reading*. London: Routledge.
Haahr, M. & S. Haahr (2010). *List randomizer* (Version 0.1) [Computer Software]. Dublin: Randomness and Integrity Services Ltd Retrieved from http://www.random.org./lists/
Horst, M. 2005. Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review* 61(3). 355–382. 10.3138/cmlr.61.3.355 (accessed 2 January 2017).
Hudson, T. 2007. *Teaching second language reading*. New York: Oxford University Press.
Huffman, J. 2014. Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language* 26(2). 17–33.
Hulstijn, J. H. 1992. Retention of inferred and given word meanings: Experiments in incidental vocabulary leaning. In P. J. Amaud & H. Bejoint (eds.), *Vocabulary and applied linguistics*, 113–125. London: Macmillan.
Ireson, J. & S. Hallam. 2001. *Ability grouping in education*. London: Sage.
Jacobs, G. M. & T. S. Farrell. 2012. *Teachers' Sourcebook for Extensive Reading*. Charlotte, NC: Information Age Publishing.
Kim, I. 2010. *Finalizing English reading I. Seoul*. South Korea: Pung-ga Kyo-yook.
Kim, Y. 2008. The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *The Modern Language Journal* 92(1). 114–130. 10.1111/j.1540-4781.2008.00690.x (accessed 2 January 2017).
Kweon, S.-O. & H.-R. Kim. 2008. Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language* 20(2). 191–215.
Laufer, B. & G.C. Ravenhorst-Kalovski. 2010. Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22 (1). 15–30.
Lee, J. & D. L. Schallert. 1997. The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly* 31(4). 713–739. 10.2307/3587757 (accessed 2 January 2017).
Millett, S., E. Quinn & P. Nation. 2007. *Asian and Pacific speed readings for ESL learners*. Wellington: English Language Institute Occasional Publication.
Milton, J. 2013. Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist & B. Laufer (eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis*, 57–78. Modena, Italy: European Second Language Association.
Nakanishi, T. 2015. A meta-analysis of extensive reading research. *TESOL Quarterly* 49(1). 6–37.
Nation, P. 2013. *What should every ESL teacher know?*. Seoul: Compass Publishing.
Nation, P. 2015. Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language* 27(1). 136–145.
Nation, P. & A. Heatley. 1996. *Vocabprofile, word and range: Programs for processing text*. Wellington, New Zealand: LALS.
Negi, J. S. (2010). *Teaching English to multilevel EFL learners in a large classroom: A pedagogical discussion*. Paper presented at the NELTA Conference, Nepal.
Paribakht, T. S. & M. Wesche. 1997. Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition: A rationale for pedagogy*, 174–200. Cambridge: Cambridge University Press.

Park, A Young. 2017. A comparison of the impact of extensive and intensive reading approaches on the Korean EFL learners' reading rate and reading comprehension development. *International Journal of Applied Linguistics & English Literature* 6(3). 131–142.
Pigada, M. & N. Schmitt. 2006. Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language* 18(1). 1–28.
Rashidi, N. & M. Piran. 2011. The effect of extensive and intensive reading on Iranian EFL learners' vocabulary size and depth. *Journal of Language Teaching and Research* 2(2). 471–482. 10.4304/jltr.2.2.471-482 (accessed 2 January 2017).
Read, J. 1993. The development of a new measure of L2 vocabulary knowledge. *Language Testing* 10(3). 355–371. 10.1177/026553229301000308 (accessed 2 January 2017).
Read, J. 2000. *Assessing vocabulary*. Cambridge: Cambridge University Press.
Renandya, W. A. & T. S. C. Farrell. 2011. "Teacher, the tape is too fast!' extensive listening in ELT. *ELT Journal* 65(1). 52–59. 10.1093/elt/ccq015 (accessed 2 January 2017).
Reynolds, B. L. 2015. The effects of word form variation and frequency on second language incidental vocabulary acquisition through reading. *Applied Linguistics Review* 6(4). 467–497.
Roche, T. & M. Harrington. 2013. Recognition vocabulary knowledge as a predictor of academic performance in an English as a foreign language setting. *Language Testing in Asia* 3(1). 1–13.
Schmitt, N., X. Jiang & W. Grabe. 2011. The percentage of words known in a text and reading comprehension. *Modern Language Journal* 95(1). 26–43. 10.1111/j.1540-4781.2011.01146.x (accessed 2 January 2017).
Senoo, Y. & K. Yonemoto. 2014. Vocabulary learning through extensive reading: A case study. *Canadian Journal of Applied Linguistics* 17(2). 1–22.
Shin, D., Y. Chon & H. Kim. 2011. Receptive and productive vocabulary sizes of high school learners: What next for the basic word list. *English Teaching* 66(3). 123–148. 10.15858/engtea.66.3.201109.123 (accessed 2 January 2017).
Song, J. & V. G. Sardegna. 2014. EFL learners' incidental acquisition of English prepositions through enhanced extensive reading instruction. *RELC Journal* 45(1). 67–84.
Stephens, M. 2016. Response to Sakurai: The influence of translation on reading amount, proficiency and speed in extensive reading. *Reading in a Foreign Language* 28(1). 151.
Tekmen, E. & A. Daloğlu. 2006. An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals* 39(2). 220–243. 10.1111/j.1944-9720.2006.tb02263.x (accessed 2 January 2017).
Treffers-Daller, J. & J. Milton. 2013. Vocabulary size revisited: The link between vocabulary size and academic achievement. *Applied Linguistics Review* 4(1). 151–172.
Vidal, K. 2011. A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning* 61(1). 219–258.
Waring, R. & M. Takaki. 2003. At what rate do learners learn and retain new vocabulary from reading a graded reader. *Reading in a Foreign Language* 15(2). 130–163.
Webb, S. 2005. Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition* 27(1). 33–52.
Webb, S. & A. C. Chang. 2015. Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning?. *Language Teaching Research* 19(6). 667–686.
Webb, S. & J. Macalister. 2013. Is text written for children useful for L2 extensive reading?. *TESOL Quarterly* 47(2). 300–322. 10.1002/tesq.70 (accessed 2 January 2017).

West, M. 1953. *A general service list of English words*. London: Longman.
Yamashita, J. 2015. In search of the nature of extensive reading in L2: Cognitive, affective, and pedagogical perspectives. *Reading in a Foreign Language* 27(1). 168–181.
Zahar, R., T. Cobb & N. Spada. 2001. Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review* 57(4). 541–572. 10.3138/cmlr.57.4.541 (accessed 2 January 2017).
Zhang, L. J. & S. B. Anual. 2008. The role of vocabulary in reading comprehension: The case of secondary school students learning English in Singapore. *RELC Journal* 39(1). 51–76.