

# The Effectiveness of Extensive Reading (ER) on the Development of EFL Learners' Sight Vocabulary Size and Reading Fluency

Akira Iwata Hokkaido Musashi Women's Junior College

#### ABSTRACT

Few studies have examined the effects of extensive reading (ER) on foreign language learners' acquisition of sight vocabulary, especially for partially known high-frequency words and reading fluency development. This study compared groups of non-English-major Japanese university-level students engaged in ER with intensive reading (IR) for two semesters. There were two ER Groups—one that read over 50,000 words (n = 21) and one that read below 50,000 words (n = 26)—and one IR Group (n = 25). The participants spent almost the same time on their task and were exposed mainly to 1k- and 2k-level vocabulary. The results of the vocabulary level tests revealed that ER was more effective and efficient for developing sight vocabulary of partially known high-frequency words. Only the ER Groups significantly improved their reading rates, indicating that ER facilitated reading fluency better than IR; moreover, the more words they read, the greater their improvement. The pedagogical implication is that the adoption of ER in an EFL curriculum is a feasible option to facilitate vocabulary acquisition and reading fluency for communication purposes in the classroom.

## **INTRODUCTION**

## **Vocabulary learning through ER**

It has been widely demonstrated that many kinds of learning occur through extensive reading (ER; Liu & Zhang, 2018; Nation & Waring, 2020). Among them, vocabulary learning and reading rate facilitation are considered among the most crucial elements for learning English as a foreign language (EFL). Vocabulary can be learned both intentionally and incidentally (Laufer, 2009). Intentional vocabulary learning is defined as word learning occurring through language-focused learning in which learners are aware that a retention test will be conducted; incidental learning is generally defined as word learning occurring as the by-product of a meaning-focused activity where learners are not aware whether a retention test will be conducted (Baddeley et al., 2009; Lindstromberg, 2020). Often, learners engage in a task involving the processing of some information through reading (Laufer & Hulstijn, 2001). Although incidental learning including ER is considered less efficient than intentional learning (Nation & Waring, 2020), it is encouraged (Webb, 2008) as a practical option (Pigada & Schmitt, 2006).

Nation and Waring (2020) categorized research on vocabulary acquisition through ER into two types: reading of the same text by all the participants (Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003) and reading of different texts (Nakanishi 2015; Pigada & Schmitt, 2006; Suk 2017). The former attempts to carefully control variables, the latter to maintain as natural an environment as possible. However, both types of research have mainly focused on learning of new vocabulary. A study by Waring and Takaki (2003) clearly demonstrated that incidental vocabulary learning occurred at several levels, namely word form recognition, multiple-choice meaning recognition, and receptive recall of the meaning. They used a graded reader and replaced 25 words of six different frequency groups with pseudo-words. They then adopted three types of test: a word form recognition test, multiple-choice meaning recognition test, and meaning recall test. These tests were administered immediately after, one week after, and three months after reading. The average scores of the immediate test were 61% for word form recognition, 42.4% for multiplechoice meaning recognition, and 18.4% for meaning recall. Scores decreased to 33.6%, 24.4%, and 3.6% after three months, respectively. Considering that the target words were pseudo-words, which would not be encountered in other texts, these results were considered to be purely the outcome of ER. The authors argued that vocabulary learning is a cumulative process and that further encounters with the words would reinforce acquisition. This research was well controlled; however, the focus was still on learning new words, and the duration of the activity was rather short (one hour). A study by Pigada and Schmitt (2006) centered on the learning of both new and partially known words. They conducted a one-month ER program with one participant, targeting 133 new and partially known words in French. They adopted a one-to-one interview test to assess knowledge of the target words' spelling, meaning, and grammatical characteristics. They showed that 49.6% of the target words were enhanced in one type of word knowledge, 9.8% in two types, and 6% in all three types, for a total of 65.4% of the target words being learned to some degree, which was quite high. This research demonstrated that repeat encounters can enhance vocabulary acquisition at least partially and that spelling knowledge can be acquired with less exposure time than is required for meaning. The interview test was informative but time-consuming, and highfrequency words showed a ceiling effect. As this was a case study of just one participant, the authors admitted that it was difficult to generalize the results.

ER using graded readers allows learners to encounter many texts with few unknown words and clear contexts (Webb, 2008). They also make the most of the repetition of high-frequency words that often appear before previous memories are forgotten (Nation & Waring, 2020; Pellicer-Sánchez & Schmitt, 2010). Researchers often adopt specially constructed measures based on the vocabulary occurring in the texts read by learners (Waring & Takaki, 2003); however, vocabulary level tests (Lee, 2007) suitable for measuring high-frequency words have also been used because they assess level mastery.

The range of high-frequency words is under debate; Schmitt and Schmitt (2012) suggested the range to be 3,000 word families, while Nation (2013) argued for 2,000 word families, which covers 86% of the British National Corpus. Regardless of the choice, it is clearly recommended that L2 learners should begin by learning and mastering high-frequency words (Nation & Newton, 1997; Şen & Kuleli, 2015).

If the ultimate goal of language learning, including vocabulary, is to enhance communication, such as face-to-face communication in real life, speed should be an important element. However, no study placed clear time limits on their measurements for individual target item, except for time limits for the entire test (e.g., Nakata et al., 2020). Therefore, this study focused on vocabulary acquisition through a vocabulary level test with a time limit.

## **Reading fluency development through ER**

Vocabulary knowledge plays an important role in EFL learners' reading processes (Laufer & Ravenhorst-Kalovski, 2010; Şen & Kuleli, 2015), especially the automatization of lower-order skills in reading such as word recognition and decoding. These lead to reading fluency, which facilitates comprehension (Grabe & Stoller, 2011). Reading fluency in this study was measured as reading speed in standard words per minute (swpm), which is defined as six adjacent characters within a text, including all letters, numbers, punctuation marks, and spaces (Carver, 1982, 1990). A number of studies have demonstrated that ER facilitates reading speed (Beglar et al., 2012; Beglar & Hunt, 2014; Bell, 2001; Bui & Macalister, 2021; Iwata, 2020; McLean & Rouault, 2017; Robb & Susser, 1989). Beglar et al. (2012) compared one intensive reading (IR) Group and three ER Groups of non-English major Japanese university students for two semesters. The approximate mean of standard words read by the IR Group was 40,000, and those of the three ER Groups were 136,000, 156,000, and 200,000. The ER Groups read mainly graded readers and un-simplified text. The authors concluded that only the ER Groups significantly improved their reading rates with adequate comprehension, and the more they read, the more they improved. Moreover, more proficient L2 learners received more benefits. The authors also argued that graded readers were more effective than un-simplified text because they contained more high-frequency words and repetition. However, the same four passages were used in the measurements, which might have produced a practice effect even after an interval of eight months. McLean and Rouault (2017) also compared an IR Group with an ER Group in a two-semester course for Japanese EFL non-English major learners. Students were randomly assigned to one of the two groups, and standard words were used to measure their reading rates. After they familiarized themselves with the timed reading procedure, three timed reading measurements in the fifth through seventh weeks were taken as pre-test results. The same texts were used at the end of the course as a part of the timed post-test. The approximate mean number of words read by the ER Group was 110,000, and that of the IR Group was 16,000. However, both groups had 15 timed reading treatments between the pre- and post-tests, which could significantly impact the development of reading speed, should be considered (Macalister, 2010). The authors concluded that both IR and ER improved reading rates with adequate comprehension; however, the ER Group showed significantly greater gains (ER: +30.96 swpm, IR: +5.15 swpm). Therefore, ER was more effective and efficient. However, three points must be considered more closely. First, the effect of timed reading could be much stronger than that of other treatments. Second, the reading rate tests in the pre- and post-tests were the same; therefore, a possible practice effect could not be excluded, even with the eight-month interval. Finally, the estimation of the amount of time spent on the activity may have not been adequately controlled.

The present author investigated the effects of ER in an earlier study with three groups, one with intensive reading plus grammar translation instruction and two with ER plus output activities, of first-year non-English major Japanese university students (Iwata, 2020). The study revealed no significant differences among the groups concerning score changes in mean reading rate. However, the first reading rate measurements were conducted not at the onset of the course but in the seventh week of the first semester, when the ER groups had already read 11,713 words on average and the participants might have already improved their reading speed. The present study builds on the earlier one, using basically the same methodology but with the following changes to permit an investigation of the effects of ER on reading fluency development: standard word units were used for the reading rate measurement; comprehension questions were adopted to ensure that adequate

comprehension accompanied the reading; however, timed reading was not included, except during the pre- and post-tests, to minimize the effect of timed reading training on reading rate development; different texts were used for the pre- and post-tests to avoid practice effects; one timed reading practice was provided prior to the pre- and post-test measurements to obviate the influence of unfamiliarity with the procedure; and a stopwatch was used for precision and to allow learners to set their own start times.

#### Purpose

The current study compared ER plus output activities with IR plus the traditional grammartranslation method (GTM) to investigate whether the former better facilitated learners' sight vocabulary learning and reading rates. A greater accumulation of research evidence is needed to convince classroom teachers and administrators that ER is comparable to traditional methods of instruction in EFL classroom settings. Hence, groups of Japanese college students engaged in ER for almost two semesters, one of which read over 50,000 words (D), the other of which read less than 50,000 words (E), and traditional IR (A), were compared. The unique point of this study is that it measured sight vocabulary under a time limit. The research questions motivating this study were as follows:

RQ1: Which vocabulary size levels significantly increased in post-treatment vocabulary level tests compared with pre-treatment tests in each group?

RQ2: Were there any differences in the changes to vocabulary size among the groups?

RQ3: Were post-treatment reading rates significantly higher than pre-treatment reading rates for the groups?

RQ4: Which group's reading rates improved the most?

## METHODOLOGY

## **Participants**

A total of 72 first-year Japanese female college students from the faculty of Economics (non-English majors) participated in this study. The students were enrolled in the only compulsory English course, General English. The participants had studied English for six years in secondary school, and their English proficiency was generally about level A1 to B1 in the Common European Framework of Reference for Languages. None of the participants had ever experienced ER instruction in English. The participants were divided into three groups based on the results of a placement test, which was administered at an institutional level, and the study was conducted using these three classes. Group A (25 students) was the most proficient, Group B (24 students) was the second most proficient, and the remaining students were placed in Group C (23 students). However, Groups B and C were reclassified according to the amount of ER completed as D (over 50,000 words) or E (under 50,000 words).

## Materials

## **Questionnaires and Interviews**

Two questionnaires were administered using Google Forms. The preliminary questionnaire, conducted in the first week of the first semester (April), covered the participants' general attitudes toward learning English and their previous English learning experiences. The final questionnaire, conducted in the 12th week of the second semester (December), covered their experiences of the course, with the addition of ER for Groups D and E. In addition to the questionnaire, weekly English class time and English study time outside the class period were reported three times (in minutes): in the 12th week of the first semester (early July), the fourth week of the second semester (early October), and the 12th week of the second semester (early December). In an attempt to obtain honest self-reports, the instructor announced that the self-reported time would not affect their final grade. An informal interview was performed during the 13th week of the second semester to confirm the responses given on the questionnaire for five to seven students from each group.

## **Vocabulary Level Tests**

The Mochizuki Vocabulary Size Test (MVST) was used in this study (Aizawa & Mochizuki, 2010) as a vocabulary level test. Although Nation's Vocabulary Levels Test (Nation, 1990) is widely used, it is not always suitable for measuring Japanese learners' vocabulary sizes (Aizawa, 1998). The MVST is widely used in a Japanese context, and is recognized to be highly reliable (Aizawa, 1998) and credible (Schmitt, Schmitt, & Clapham, 2001). There are three sets of this test that are designed to be equivalent levels, and each test consists of items from the 1k-word level to the 7k-word level. Each level contains 25 items, and test takers must choose a word with the meaning equivalent to that of the first language (L1) item shown on the computer screen from among three target word options within five seconds. The pre-test was conducted in the first week of the first semester, and the post-test was conducted in the 12th week of the second semester. Different sets of tests were used each time.

#### **Reading Rate Tests**

The features of the reading rate tests were as follows. Three different texts (A, B, and C) were the same ones used in the previous research (Iwata, 2020). However, Texts B and C in this research were reversed from the previous study in order to minimize the influence of different background knowledge on the topic of the participants. Text A was used for the pre-test and Text B was used for the post-test, while Text C was used for practice before tackling Texts A and B. They were all narratives, and the topics were School Life (A & B) and Family (C). Each test had five true or false fact-finding comprehension questions to check whether the participants sacrificed comprehension for speed. The total standard word units were 149.33, 129.83, and 131.50, respectively. The readability of each text and the 1k- and 2k-word level vocabulary coverage are shown in Table 1. Although the homogeneity of the three texts was confirmed in the previous study, additional reading rate tests were conducted for the two texts used as pretest and posttest. A total of 47 English major students who did not participate in this study read text C first and then the other two texts A and B in a different order and answered comprehension questions for each.

Table 2 shows the average reading rate for each text. A paired-sample *t*-test was conducted on the standard words per minute (swpm) for these texts, confirming that there was no significant difference among them with almost no effect size, t (46) = 0.786, p = 0.436, r = .005.

Table 1. Readability of Each Text							
Text	Total Number of Words (Standard word unit)	Flesh Kincaid Ease	Flesh Kincaid Grade Level	1k- and 2k-word Level Vocabulary Coverage (Percentage)			
A (Pre)	149.33	73.6	6.0	93.76			
B (Post)	129.83	74.3	6.3	96.55			

**Table 1.** Readability of Each Text

Table 2. Mean Wpm for Each Text (Standard Word Units)						
Text	A (P	re)	B (Post)			
	М	SD	M	SD		

25.51

110.62

108.76

24.35

#### **Graded Readers**

N = 47

Approximately 2,500 graded readers were available in the college library, 65% of which were from levels zero through three in the M-Reader system. Table 3 shows the readability of the books randomly chosen from each level. This indicates that the participants had a wide selection of beginner-level books they could borrow whenever the library was open, which is crucial for an ER course (Day & Bamford, 2002). The proportion of correct answers on the M-Reader comprehension quizzes was set as 60%, which had served adequately in the previous study. Anderson (2008) argued that understanding 70% of a text indicates adequate comprehension; however, the non-English major students, especially those whose reading proficiency levels were not high enough, often struggled to comprehend the questions themselves. Thus, lowering the minimum threshold could help them avoid this problem and the demotivation due to successive unsuccessful attempts.

Level (M- Reader)	Total Number of Words	Flesh Kincaid Ease	Flesh Kincaid Grade Level	1k- and 2k-word Level Vocabulary Coverage (Percentage)				
0	71	97.4	0.7	78.67				
1	443	95.5	1.3	84.66				
2	890	95.3	1.8	87.62				
3	1,243	88.7	2.6	92.85				

Table 3. Readability of Sample Books

## **Instructional Method**

As in the previous research (Iwata, 2020), the course lasted for two semesters consisting of 15 weeks each, and classes of 90 minutes were conducted once a week. Two different teachers taught the three groups, both of whom were experts in English grammar and had over 30 years of

teaching experience. Group A was taught mainly by the IR method, with a focus on GTM. Each unit of the textbook consisted of an approximately 304.3-average-word text with vocabulary explanation followed by an approximately 176.2-average-word dialogue with composition drills focusing on particular grammatical points and formulaic expressions. The average 304.3-word texts were narratives explaining famous cities and places such as Paris, New York, and San Francisco. The average readability of the texts was 65.9 (Flesh Kincaid Ease; FKE) and 8.5 (Flesh Kincaid Grade Level; FKGL). The average 176.2-word dialogues dealt with situations such as a hotel front desk, the airport, and a restaurant. The average readability of the dialogues was 90.3 (FKE) and 2.3 (FKGL). Vocabulary frequency in the average 304.3-word texts and the 176.2-word dialogues covered 88.1% and 95.8% of the 1k- and 2k-word levels (high frequency) on average, according to VocabProfilers (Cobb, 2018). The class had covered 13 units by the time the final tests were conducted, representing at least 6,247 words, excluding rereading. Groups B and C (D and E) were taught by the same teacher using identical procedures. The first 30 minutes were spent on in-class ER, and the subsequent 10 minutes were spent on a book report session. The rest of the class time was spent on speaking activities using a textbook that included approximately 210.9word texts consisting of three model speech texts, introducing Japanese culture and life. The texts were narratives whose average readability was 86.3 (FKE) and 2.8 (FKGL). Vocabulary frequency at the 1k- and 2k-word levels covered 97.41%. The class went over 29 units and the total number of words read in the textbook was 6,117 words. They were encouraged to read as much as possible in and outside the class. The M-Reader system was adopted for the word count and a total of 50,000 words was set as the minimum requirement for the academic year, making up 20% of the final grade. Additional points were given for additional words read for an extra incentive.

## Procedure

A general orientation to the course was conducted in the first week of the first semester, followed by an explanation of the procedure for the reading rate pre-test by the author. All the participants tackled Text C as practice while timing themselves using their smartphones. They flipped the page and answered five comprehension questions without returning to the text. They then followed the same procedure for Text A. After the tests were collected, they tackled the first MVST and the preliminary questionnaire. Group A returned to their classroom and had normal class instruction, while Groups B and C received an orientation to ER. Normal class instruction progressed from the 2nd week to the 14th week, followed by a test week and a month of summer vacation. Weekly English study time outside the class period was checked in the 12th week (early July). There was no ER during the vacation or the 15th week. The second semester commenced right after the 15th week in September with normal instruction continuing until the 11th week in December. Weekly English study time was checked in the fourth week (early October). The reading rate post-test, post-MVST, and third weekly English study time check were conducted in the 12th week (early December). At the end, the ER data were obtained from the M-Reader system, and the final questionnaire was administered. An informal interview was performed during the 13th week.

## **Data Collection and Analysis**

This research was supported in part by the Japan Society for the Promotion of Science (JSPS) and was conducted in accordance with the ethical policy of the JSPS (2015). Prior to any

data collection, all the participants were notified of the following: This is a research project on English classes, the participation would be voluntary, all the data collected in the course would be treated as anonymous and confidential, and the data would not affect the final grade of the course. After this, the informed consent form was delivered and all the participants signed it.

The data from the MVST and reading rate tests for Group A (IR), Group D (more than 50,000 words), Group E (less than 50,000 words), and Group D+E (ER) were compared. The scores of the vocabulary level pre- and post-tests were compared using paired sample *t*-tests for each level. To compare the effects of different instruction methods, an analysis of covariance (ANCOVA) was conducted for the vocabulary levels of 1k, 2k, and 3k. The analysis of the remaining vocabulary levels (4k, 5k, 6k, and 7k) was omitted because little vocabulary from these levels was included in the reading materials they read. As a post-hoc test, a Bonferroni test was conducted for the lower vocabulary levels. The results of the reading rate pre- and post-tests (swpm) were compared using paired-sample t-tests. To compare the effects of the different instruction methods, ANOVA and paired-sample *t*-tests were conducted on the increments of the three groups (A, D, and E) and between Groups A (IR) and D+E (ER), respectively. As a post-hoc test, a Bonferroni and a Ryan's method test were adopted. The average of the three surveys (July, October, and December) on weekly English study time outside the class period were compared among Groups A, D, and E as well as between A (IR) and D+E (ER) using an ANOVA and pairedsample *t*-tests, respectively. As a post-hoc test, Ryan's method was adopted. The results of the preliminary and final questionnaires, which were scored on a five-point Likert scale, free written responses, and an informal interview were also compared.

#### RESULTS

The preliminary questionnaire (Table 4) confirmed that the mean scores of all the items except one were below three points, meaning that all the Groups, especially E, were not confident about their English skills.

	Group	D ( <i>n</i> = 21)	E ( <i>n</i> = 26)	A ( <i>n</i> = 25)
Item	Score*	М	M	М
Do you like studying English?		2.90	2.50	2.88
Are you good at English?		2.05	1.62	2.24
Do you like reading English books?		2.71	2.23	2.84
Are you confident about reading English?		2.05	1.85	2.08
Do you like reading Japanese books?		3.00	3.46	3.64

Table 4. The Preliminary Questionnaire

**Note**: \* 5-point Likert scale (5: *Strongly agree*, 4: *Somewhat agree*, 3: *Neutral*, 2: *Somewhat disagree*, 1: *Strongly disagree*)

Table 5 shows the number of words read by Groups D and E. Table 6 shows the average number of quizzes D and E took (the number of books they read), passed, and failed. If they failed a book, both the number of words and the books were excluded from the count. The passing rates were 92.72% for D and 88.29% for E based on the criterion of 6/10 correct answers. Table 7 shows

the percentage of books the participants read at each level, 91.38% of which belonged to levels 0-2.

<b>Table 5.</b> Number of Words Read by Groups D and E								
	M	SD	Max	Min				
D ( <i>n</i> = 21)	77,384.57	30,586.61	190,799.00	50,577.00				
E ( <i>n</i> = 26)	36,965.81	7,439.47	49,867.00	23,487.00				
Total $(n = 47)$	55,025.26	29,144.66						

	Taken		Passed		Failed		
	М	SD	М	SD	М	SD	<b>Passing Rate</b>
D ( <i>n</i> = 21)	108.52	56.42	100.62	54.31	7.90	5.75	92.72%
E ( <i>n</i> = 26)	58.12	21.51	51.31	18.96	6.81	6.62	88.29%
Total $(n = 47)$	80.64	47.72	73.34	45.74	7.30	6.20	90.95%

Table 6. Number of Quizzes Taken by Groups D and E

**Table 7.** Percentage of Books Read by Participants at each Level

Level (M-Reader)	Number of books	Percentage
0	1051	30.49
1	1136	32.96
2	963	27.94
3	216	6.27
4	79	2.29
5	1	0.03
6	1	0.03
Total	3447	100

Paired-sample *t*-tests were conducted on the scores of the pre- and post-MVST as vocabulary level tests for 1k to 3k. Table 8 shows the results of each instructional method and conditions, while Figures 1, 2, and 3 highlight the mean scores of 1k, 2k, and 3k, respectively. Significant differences were found for Groups D, E, and D+E (ER) in 1k; all the Groups in 2k; D, E, and D+E (ER) in 3k.

An ANCOVA was conducted on the scores of the MVST for both the pre- and post-tests for the levels from 1k to 3k, with the instructional method (IR vs. ER over 50,000 vs. ER under 50,000) used as a between-subjects independent variable for the groups with different average vocabulary sizes. The results of the pre-tests were covariates. The same ANCOVA was conducted between A (IR) and D+E (ER). For the 1k level, the ANCOVA revealed no significant differences among the groups with a medium effect size, F(2, 68) = 2.553, p = .085, partial  $\eta^2 = .070$ . Moreover, no significant difference was found between Groups A (IR) and D+E (ER) with almost no effect size, F(1, 69) = 0.148, p = .702,  $\eta^2 = .002$ . A ceiling effect (Shimizu et al., 2013), in which the mean score plus standard deviation exceeded the maximum score, was found in all groups with a medium effect size, F(2, 68) = 4.634, p = .013, partial  $\eta^2 = .120$ . A Bonferroni post hoc test was conducted, showing significant differences between A and D and between A and E, with a calculated alpha set at .0167. Moreover, a significant difference was found between A (IR) and D+E (ER) with a medium effect size, F(1, 69) = 8.949, p = .004,  $\eta^2 = .115$ . For the 3k level, the ANCOVA revealed no significant difference among the groups with a medium effect size, F(2, 68) = 2.196, p = .119, partial  $\eta^2 = .061$ . Moreover, no significant difference was found between A (IR) and D+E (ER) with a small effect size, F(1, 69) = 3.482, p = .066,  $\eta^2 = .048$ .

Level	Group	Pre	Post	Difference	t	<i>p</i> (two-tailed)	Effe	ct Size ( r )
	A (IR)	987.20 (29.93)	976.00 (25.82)	-11.20	1.77	.090	.35	Medium
	D	960.00 (43.82)	980.95 (34.91)	20.95	-2.33	.030	.51	Large
1k	Е	(43.82) 932.31 (61.73)	(34.91) 955.38 (32.65)	23.07	-2.12	.045	.41	Medium
	D + E (ER)	(01.73) 944.68 (55.67)	(32.03) 966.81 (35.70)	22.13	-3.08	.003	.45	Medium
	A (IR)	860.80 (107.16)	913.60 (73.65)	52.80	-3.01	.006	.60	Large
-	D	683.81 (114.48)	761.90 (100.78)	78.09	-3.15	.005	.69	Large
2k	Е	649.23 (127.15)	730.77 (115.69)	81.54	-3.67	.001	.72	Large
	D+E (ER)	664.68 (121.60)	744.68 (109.24)	80.00	-4.89	.000	.71	Large
	A (IR)	718.40 (89.05)	684.80 (148.02)	-33.60	1.49	.150	.30	medium
	D	628.57 (117.40)	571.43 (140.37)	-57.14	3.21	.004	.70	Large
эк	Ε	575.38 (107.97)	498.46 (119.32)	-76.92	3.08	.005	.60	large
	D+E (ER)	599.15 (114.20)	531.06 (132.85)	-68.09	4.29	.000	.63	large

**Table 8.** The Results of The Vocabulary Level Test 1k to 3k



Figure 1. Results of the vocabulary level tests (1k)



Figure 2. Results of the vocabulary level tests (2k)



Figure 3. Results of the vocabulary level tests (3k)

Paired-sample *t*-tests were conducted on the scores of the reading rate pre- and post-tests. Table 9 shows the results for different conditions, while Figure 4 highlights the mean scores. The results indicated that ER Groups (D, E, and D+E) showed significant improvements with large effects sizes, but while Group A showed improvements, they were not significant. An ANOVA was conducted on the differences between the pre- and post-test measurements for the three groups, and a significant difference was found among the groups with a medium effect size, F(2, 69) = 5.385, p = .007,  $\eta^2 = .135$ . A Bonferroni post hoc test was conducted, showing significant differences between D and E, with a calculated alpha of .0167. A *t*-test was also conducted between Groups A (IR) and D+E (ER), and the result revealed that D+E (ER) outperformed A (IR) with a large effect size, t(70) = -2.09, p = .040, r = .55.

Group	Pre	Post	Difference	t	p(two-tailed)	Effe	et Size (r)
$\Lambda$ (IP)	79.98	83.61	3 63	-2.02	054	/1	Medium
A (IK)	(23.13)	(22.67)	5.05	2.02	.034	.41	Wicdfulli
Л	79.71	95.15	15 44	_4.40	000	06	Largo
D	(22.43)	(12.75)	15.44	4.40	.000	.90	Laige
F	73.11	79.42	631	-2 58	016	50	Larga
Ľ	(19.60)	(22.02)	0.31	-2.38	.010	.30	Large
D+E (ED)	76.06	86.44	10.38	_1 82	000	70	Largo
D+E (ER)	(21.40)	(20.27)	10.38	4.02	.000	.70	Large

Table 9. Results of the Reading Rate Tests



Figure 4. Results of the reading rate tests

Table 10 shows the mean and standard deviation of weekly English studying time (minute) at three measurements ( $\alpha = .84$ ) for the participants. An ANOVA was conducted on the mean weekly study time for the groups as a between-subjects variable. The results revealed a significant difference among the groups with a large effect size, F(2, 69) = 14.855, p = .000004,  $\eta^2 = .30$ . A Bonferroni post hoc test was conducted, showing a significant difference between A and D, A and E, and D and E with a calculated alpha of .0167. However, the result of the *t*-test on the mean weekly study time between Groups A (IR) and D+E (ER) showed no significant difference with almost no effect size, t(70) = 0.173, p = .865, d = .04. Table 11 shows the results of the final questionnaire.

Table 10. Results of the Mean Weekly Study Time (minute)

		J		)
Group	July	Oct	Dec	М
A (IR)	278.40	266.80	274.80	273.33
<i>n</i> = 25	(75.43)	(71.63)	(74.06)	(51.92)
D (ER-H)	333.81	302.86	312.86	316.51
<i>n</i> = 21	(54.27)	(56.40)	(65.81)	(53.87)
E (ER-L)	227.69	229.23	235.00	230.64
<i>n</i> = 26	(59.69)	(58.10)	(61.53)	(55.47)
D + E (ER)	275.11	262.13	269.79	269.01
<i>n</i> = 47	(77.85)	(67.73)	(73.97)	(69.25)

•	<b>Fable 1</b>	1. The	Final	Question	naire
---	----------------	--------	-------	----------	-------

	Group	D ( <i>n</i> = 21)	E ( <i>n</i> = 26)	A ( <i>n</i> = 25)
Item	Score*	М	М	М
Extensive reading was interesting.		3.43	2.73	
I felt fulfilled after I finished reading books.		4.10	3.73	

My vocabulary knowledge improved through reading English.	3.48	2.81	3.64
I became able to infer the meaning of unknown words from context.	3.81	3.23	2.88
The meanings of words come to my mind without trying to do so.	3.10	2.81	2.76
I looked up unknown words in the dictionary while reading.	2.67	2.92	4.24

Note: \* 5-point Likert scale (5: *Strongly agree*, 4: *Somewhat agree*, 3: *Neutral*, 2: *Somewhat disagree*, 1: *Strongly disagree*)

#### DISCUSSION

RQ1 asked about the vocabulary size levels that became significantly larger in the posttreatment vocabulary level tests. Group A (IR) read approximately 6,000 words, excluding repetition, 92% of which were at the 1k and 2k levels. Group D+E (ER) read 55,025.26 words, on average, in graded readers (D: 77,384.57, E: 36,965.81), plus 6,117 words from the textbook. Over 90% of the graded readers they read were levels 0 to 2. The 2k-level vocabulary coverage was 84% for the graded readers and 97% for the textbook. The scores of all the groups showed a ceiling effect for the 1k level and significantly increased for the 2k level. The scores for the 3k level decreased significantly. The 3k level comprised vocabulary they rarely encountered. Considering that even Group A improved at the 2k level, a moderate number of encounters with words is still significant for high-frequency words to become sight vocabulary.

RQ2 asked whether there were differences in the changes to each level of vocabulary size among the groups. There were no significant differences between any groups for 1k; however, the ceiling effect was found in all groups. The reason that Groups D and E showed a greater difference is likely because Group A started off closer to the highest possible score of 1,000, which allowed them less room for improvement. There were significant differences between A and D, A and E, and A (IR) and D+E (ER), indicating that ER using graded readers was more effective for the acquisition of sight vocabulary than IR, likely because of the greater amount of repetition (Beglar et al., 2012; Nation & Waring, 2020). Given that there was no significant difference in study time between A (IR) and D+E (ER), ER was more effective and efficient for incidental sight vocabulary learning of high-frequency words that were partially familiar. However, there were significant differences in study time among the three Groups (D > A > E). Group E spent the least time on study, perhaps because they were less motivated, as revealed by the preliminary questionnaire, but still showed improvement. More research is needed to determine whether learners acquire more words when they read more.

RQ3 asked whether the post-treatment reading rates were significantly higher than the pretreatment reading rates for each group. Group A (IR) improved their reading rate with a medium effect size, but it was not significant, while D, E, and D+E (ER) significantly improved their reading rates with large effect sizes. Repetition of vocabulary could help in the acquisition of sight vocabulary, as it reduces the cognitive load during lower-order skills such as word recognition or decoding (Grabe & Stoller, 2011) and results in faster reading. This study did not adopt timed reading, and there were only four opportunities for practice and testing. Therefore, the different results were derived purely from the differences between IR and ER. These results are in line with previous research (Beglar et al., 2012; McLean & Rouault, 2017).

RQ4 asked which group improved their reading rates the most. Group D improved their reading rates more significantly than A and E, which was in line with previous research (e.g., McLean & Rouault, 2017). Again, there was no significant difference in study time between A (IR) and D+E (ER); therefore, ER was more effective and efficient for facilitating reading fluency. In addition, the results seem to support previous research findings that the more students read, the faster they read (Beglar et al., 2012).

The final questionnaire revealed no large differences in their feelings about vocabulary learning. The ER Group improved more in the 2k vocabulary, despite the fact that Group A reported using a dictionary far more often, which could have facilitated vocabulary learning (Laufer, 2003). The questionnaire and interviews revealed that most of the ER Group participants felt rather confident in passing the quizzes while roughly following the story and not using the dictionary often. However, there were a few learners in Group E who said that they found it hard to understand the questions in the M-Reader quizzes and felt pressured by the time limit. In this respect, 60% is appropriate for learners who are not very proficient. Overall, the ER participants had a positive view of the experience.

#### **Pedagogical Implications**

Learning high-frequency words should be the first step for EFL learners because they cover a very large proportion of words appearing in spoken and written texts in formal and informal uses of the language. The more they learn, the larger the variety of activities they can do in the target language. Extensive reading can enable learners to repeatedly encounter high-frequency words in a variety of contexts, which would better enhance their understanding than merely learning their L1 equivalents. When the environment is right, extensive reading can also offer an ideal learning environment where learners can receive extensive exposure to the target language text with less suffering and struggle even if they do not like studying the target language very much or are not confident about their ability in the language, because they have freedom to choose the book they read depending on their ability and interest. This means that extensive reading is feasible no matter how large the gap in the levels of language proficiency among learners in a class, and there is virtually no class without differences in learners' proficiency and interest. Offered an appropriate ER program, learners, including ones who are not highly motivated or confident in learning English, would hopefully feel joy, self-efficacy, and fulfillment as in this study because they realize that they can read in the target language, which will be a priceless experience for them. Regardless of whether the program adopts real books or online books, it will require a relatively high cost in the beginning. However, considering its effectiveness, it will definitely be worth the cost. Therefore, ER should be adopted in EFL classrooms along with intentional vocabulary learning.

#### Limitations

This study was conducted as part of normal class instruction and reflects the reality of working with intact classes. The most proficient group could have benefited the most if they engaged in ER (Beglar et al., 2012). In addition, it was difficult to control the time on task, as the more the participants read, the more time they spent, and how much they read depended on individual motivation or proficiency.

## **CONCLUSION AND FUTURE RESEARCH**

This study investigated the impact of ER on the acquisition of sight vocabulary, including partially known items and the development of reading fluency, based on a comparison with a traditional instructional methodology. The ER and IR Groups spent almost the same time on studying English, given the participants' self-report of their study time, showing that ER was a more effective and efficient method for these two areas of learning. The ER approach in this study is feasible for almost any classroom and is an effective and efficient way of reinforcing vocabulary and developing reading fluency, even for learners who are not highly motivated or proficient, as it allows them to enjoy reading and to feel fulfilled. It is recommended that ER be adopted as part of a balanced learning program (Nation & Waring, 2020). Future research will shed light on how ER facilitates gains in sight vocabulary together with intentional learning.

#### ACKNOWLEDGEMENT

This research was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Number JP18K00899)

#### REFERENCES

- Aizawa, K. (1998). Developing a vocabulary size test for Japanese EFL learners. *Annual Review* of English Language Education in Japan, 9, 75-85.
- Aizawa, K., & Mochizuki, M. (2010). *Practical handbook for English vocabulary teaching*. Taishukanshoten.
- Anderson, N. J., & Nunan, D. (2008). *Practical English language teaching: Reading*. McGraw-Hill.
- Baddeley, A., Eysenck, M., & Anderson, M. (2009). Memory. Psychology Press.
- Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language*, *26*, 29-48. https://nflrc.hawaii.edu/rfl/item/290
- Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning*, *62*, 665-703. https://doi.org/10.1111/j.1467-9922.2011.00651.x
- Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix*, *1*(1). http://www.readingmatrix.com/articles/bell/index.html
- Bui, T. N., & Macalister, J. (2021). Online extensive reading in an EFL context: Investigation reading fluency and perceptions. *Reading in a Foreign Language*, 33(1), 1-29. https://nflrc.hawaii.edu/rfl/item/523
- Carver, R. P. (1982). Optimal rate of reading prose. *Reading Research Quarterly*, *18*, 56-88. https://doi.org/10.2307/747538.
- Carver, R. P. (1990). Reading rate: A review of research and theory. Academic Press.
- Day, R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, *14*(2), 136-141. https://nflrc.hawaii.edu/rfl/item/61
- Grabe, W., & Stoller, F. L. (2011). Teaching and researching reading (2nd ed.). Pearson.

- Iwata, A. (2020). How extensively do we need to read to improve EFL reading ability? A comparison of two different instructional methodologies. *The Reading Matrix*, 20(1), 66-83. http://www.readingmatrix.com/files/22-15971381.pdf
- Japan Society for the Promotion of Science. (2015, February). For the sound development of science: The attitude of a conscientious scientist. Japan Society for the Promotion of Science. https://www.jsps.go.jp/j-kousei/rinri.html
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review*, 59(4), 567-587. https://doi.org/10.3138/cmlr.59.4.567
- Laufer, B. (2009). Second language vocabulary acquisition from language input and from form focused activities. *Language Teaching*, 42, 341-354. https://doi.org/10.1017/S0261444809005771
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26. http://dx.doi.org/10.1093/applin/22.1.1
- 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), 15-30. https://nflrc.hawaii.edu/rfl/item/206
- Lee, S. Y. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal*, 38(2), 150-170. https://doi.org/10.1177/0033688207079730
- Lindstromberg, S. (2020). Intentional L2 vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 240-254). Routledge.
- Liu, J., & Zhang, J. (2018). The effects of extensive reading on English vocabulary learning: A meta-analysis. *English Language Teaching*, 11(6), 1-15. https://doi.org/10.5539/elt.v1n6p1
- Macalister, J. (2010). Speed reading courses and their effect on reading authentic texts: A preliminary investigation. *Reading in a Foreign Language*, 22(1), 104-116. https://nflrc.hawaii.edu/rfl/item/211
- McLean, S., & Rouault, G. (2017). The effectiveness and efficiency of extensive reading at developing reading rates. *System*, 70(1), 92-106. https://doi.org/10.1016/j.system.2017.09.003
- Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly*, 49(1), 6-37. https://doi.org/10.1002/tesq.157
- Nakata, T., Tada, S., McLean, S., & Kim, Y. A. (2020). Effects of distributed retrieval practice over a semester: Cumulative tests as a way to facilitate second language vocabulary learning. *TESOL Quarterly*, 55(1), 248-270. https://doi.org/10.1002/tesq.596
- Nation, I. S. P. (1990). Teaching and learning vocabulary. Heinle & Heinle Publishers.
- Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Heinle, Cengage Learning.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Nation, P., & Newton, J. (1997). Teaching vocabulary. In J. Coady & T. Huckin, (Eds.) *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 238-254). Cambridge University Press.
- Nation, I. S. P., & Waring, R. (2020). *Teaching extensive reading in another language*. Routledge.

- Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, 22(1), 31-55. https://nflrc.hawaii.edu/rfl/item/207
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1), 1-28. https://nflrc.hawaii.edu/rfl/item/114
- Robb, T. N. & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, 5(2), 239-251. https://nflrc.hawaii.edu/rfl/item/471
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329-363. https://doi.org/10.1177/1362168808089921
- Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, doi:10.1017/S0261444812000018.
- Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, 18, 55-88. https://doi.org/10.1177/026553220101800103
- Şen, Y., & Kuleli, M. (2015). The effect of vocabulary size and vocabulary depth on reading in EFL context. *Procedia—Social and Behavioral Sciences*, 199, 555-562. https://doi.org/10.1016/j.sbspro.2015.07.546
- Shimizu, K., Miho, N., & Yamamoto, R. (2013). Extension of the common factor space to new variables or variables included in parcels: A comparison among estimates by factor score method and Gorsuch's method. *Kansai Daigaku Shinrigaku Kenkyu*, 4, 33-39.
- Suk, N. (2017). The effects of extensive reading on reading comprehension, reading rate, and vocabulary acquisition. *Reading Research Quarterly*, *52(1)*, 73-89. https://doi.org/10.1002/rrq.152
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130-163. https://nflrc.hawaii.edu/rfl/item/78
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232-245. https://nflrc.hawaii.edu/rfl/item/178

Akira Iwata is a Professor at Hokkaido Musashi Women's Junior College, Japan. He holds a Master of Education in TESOL from the University of Wollongong, Australia. His research interests include EFL classroom practices, particularly EFL pedagogy, such as reading, writing, and vocabulary learning instructions.

Email: iwata@hmjc.ac.jp