How Extensively Do We Need to Read to Improve EFL Reading Ability?: A Comparison of Two Different Instructional Methodologies

Akira Iwata
Hokkaido Musashi Women's Junior College

ABSTRACT

This study compares two instructional methodologies: extensive reading plus output activity and intensive reading plus grammar-translation with regard to improvement of non-English major EFL learners' reading comprehension and fluency development in a Japanese junior college. It identifies the minimum number of words an individual should read for improvement, developing a classroom instructional model that can be implemented as a national policy. The results indicated that 50,000 words should be read to be as effective as conventional instructional methodology. Reading fluency can be improved regardless of instruction types or words read. The study time could peak before tertiary study, aligning with extrinsic motivation to study English. The results of questionnaires enquiring about their attitudes towards English learning and extensive reading are also discussed.

INTRODUCTION

Background

The effectiveness of Extensive Reading is now recognized, not only for an individual's first language, but also for developing a second or foreign language, especially within reading pedagogy (Day & Bamford, 1998). This process covers various skills and aspects of language development, from building motivation levels to improving writing skills (e.g. Park, 2015). Extensive reading has even been claimed to be the second most effective method for improving target language ability, behind living in a native-speaking country of the language in question (Nuttall, 1996), or, alternatively, as the single most effective approach for improving target language proficiency (Malay, 2005). Indeed, no one would consider that being able to read extensively in a foreign language would impede learners' development of their target language. Since the 1990s, extensive reading has gained increasing levels of interest, not only in ESL settings, but also in EFL settings, which are considered to be more effective than the former. Despite the fact that this method is now widely recognized by many practitioners, it has yet to spread into classrooms both within ESL and EFL courses (Grabe, 2009; Jeon & Day, 2016). One explanation for this may be due to many teachers and administrators asserting that there is not enough evidence to validate the benefits of Extensive Reading (Huffman, 2014). Therefore, it is vital to accumulate research that deals with this particular issue, including whether such a practice

is conducted in ESL or EFL, the respective countries' educational policies, and the participants' ages etc. This research focuses on the context of a Japanese women's junior college.

The Ministry of Education, Culture, Sports, Science and Technology (2017) reports that 80.6% of Japanese students are enrolled in tertiary education. In terms of the general course of secondary education, the curriculum is set by the government according to the course of study and, consequently, there is relatively little difference between the number of English classes. However, in tertiary education, the number of English classes varies greatly. Furthermore, students who major in English-related courses are more likely to study the language spontaneously. Given this state of affairs, research into non-English-related-course-major students would more accurately reflect the situation of the majority of Japanese students.

Also, it is widely known that grammar-translation methods remain the most popular instructional methodology in Japan, especially in secondary education (Kanatani, Takayama, Usukura, & Ota, 2011), so this situation is likely to be very similar in the context of tertiary education. In order to change this state of affairs, raising awareness of the effectiveness of extensive reading and convincing teachers and administrators of that fact should be prioritized, as suggested by Macalister (2010). Generally speaking, Japanese teachers tend not to be in favor of sudden and drastic changes to their teaching methods. Therefore, as Bell (2001) argues, comparing extensive reading with conventional classroom teaching methods is required to empirically demonstrate the effectiveness of this method. Furthermore, it is also vital to outline possible models for instruction regarding effective implementation of this approach (Jeon & Day, 2015). Showing how many words students need to read in order to improve their reading ability forms an important aspect of this task.

How extensively should we read to improve reading speed and comprehension?

Reading is often regarded as the most important language skill in EFL academic contexts (Grabe, 1991), and this is borne out by the fact that most English textbooks for secondary education approved by the Ministry of Education, Culture, Sports, Science and Technology in Japan contain reading text as the main part of each unit. The two most important perspectives used to indicate learners' reading ability should be comprehension and fluency. This is because comprehension should be accompanied by reading fluency; meaning that, in order to comprehend a text, one needs to be able to read quickly (Breznits, 1988; Huffman, 2014). If readers can only process information at a slow rate, and with limited attentional resources, they will be unable to maintain enough detail in their short-term memory to decode the overall message of the text (Brown & Hirst, 1983). Therefore, extensive reading offers an ideal solution to this problem because, as Grabe and Stoller (2002) argue, students need to engage in reading for a long time with material that is set at an appropriate level for their ability. It is true that an abundant amount of research has attempted to highlight the effects of extensive reading on the improvement of language proficiency. However, relatively little research has been conducted regarding L2 fluency development (Beglar, Hunt, & Kite, 2012; Grabe, 2009) and reading comprehension.

There is research that investigated the effectiveness of extensive reading on developing reading fluency and reading comprehension. However, the research often contains some shortcomings. Some asserted that the participants' reading fluency, but not comprehension, improved. However, they often lack information on how extensively they read (Fujita & Noro, 2009; Lai, 1993; Matsui & Noro, 2010; Yamashita, 2008). Other research suggested the improvement of both reading fluency and comprehension. However, they often had limitations,

including very short course duration, or the tests used to measure reading rates were not reliable enough, or no control group existed (Iwahori, 2008; Nakanishi & Ueda, 2011; Lai, 1993). Still other studies that have attempted to compare a range of tasks with extensive reading had drawbacks. Robb and Susser (1989), for example, compared extensive reading with skill building in the context of Japanese EFL college students, finding that the former is as effective as the latter. The main strength of this study was that it compared extensive reading with a traditional teaching method. However, the extent to which the participants actually read was unclear.

Bell (2001) compared extensive reading and intensive reading approaches for young adult students in Yemen, focusing on the development of their reading speeds and comprehension across two semesters. Graded readers were used for extensive reading, and the results suggested that the extensive reading group significantly improved both their reading speeds and their comprehension. However, the total number of subjects included in the study was only 26, meaning that individual differences could impact the results dramatically.

Beglar, Hunt, and Kite (2012) compared extensive reading and intensive reading methods for improvements in reading fluency and comprehension for Japanese first year university students over a one-year period. They concluded that the only extensive reading group significantly improved their reading rates, while also finding that the group that read the most made greater improvements than the group that read the least. Also, unlike L1 reading, lower proficiency readers do not always benefit from this approach. Instead, the amount of reading is likely to be associated more strongly with levels of fluency. They also argue that using simplified texts, such as graded readers, offers a more beneficial method. However, we should be careful about three points to better understand the research. First, given the fact that the participants could read unsimplified books like Bridget Jones's Diary as well as high level graded readers (Oxford level 5 to 6), they are highly proficient readers with very high motivation. Second, the teachers selected the six graded readers, not the participants. Therefore, it could be argued that the books were not those preferred by the participants and so did not meet a requirement of Extensive Reading where learners choose what they want to read (Day and Bamford, 2002). Finally, as the researchers admitted that the same reading rate test was used for pre- and post-test. Even though the answers to the comprehension questions were not shown to the participants, they could have checked some vocabulary, expressions, or grammar after the test because they were highly motivated to learn English.

In terms of the reading quantity, Beglar and Hunt (2014) have reported how simplified texts are more effective than unsimplified texts in the context of 14 Japanese students who improved their reading rate significantly by reading over 200,000 words. The amount of words the participants in this study who developed their reading rate ranges from 12,000 to 330,000 words in a course of under one year, and from 136,000 to 400,000 words for longer courses. Nation (2009) has recommended 500,000 words a year as a benchmark. However, as Beglar and Hunt (2014) argue, reading such an amount in a single year may not be realistic. Through his research on non-English major Japanese university students, Hagley (2017) argues that 85,000 words per semester (170,000 words a year) should constitute a realistic target that can guarantee improvement. As he suggests, students are very busy and competition for their time is fierce. This issue represents the main reason why students cannot read extensively (Ro, 2014), and the situation may be even worse in junior college due to the pressures of job hunting, which usually starts from the first year. In this situation, it is important to allocate time in class for the purpose of extensive reading, as well as encouraging students to read outside the classroom (Jeon & Day, 2016). This

is because students will most certainly not read unless it is implemented as a part of the curriculum and is properly evaluated (Hagley, 2017).

Finally, it is frequently pointed out that extensive reading should be incorporated into the curriculum and integrated with other classroom-based activities. Therefore, extensive reading plus output activities will be compared with intensive reading plus traditional grammar-translation methods. If this approach works well, it can help to convince teachers of its viability as a possible model. The reading comprehension test should be developed using reliable high stakes tests, while the reading rate test should be designed at an appropriate level for the participants using comprehension questions. The number of words and the duration of the course should also be realistic.

Study Purpose

The purpose of this research is to investigate whether extensive reading plus output activities can better facilitate participants' reading comprehension and reading fluency than traditional grammar-translation methods plus intensive reading. It also attempts to reveal how many words need to be read to make a noticeable difference, as well as to prescribe a possible instructional model for Japanese EFL classrooms. In pursuit of these goals, the following research questions and hypothesis were set:

RQ1: Does extensive reading plus an output activity better facilitate learners' reading comprehension and reading fluency than traditional intensive reading plus grammar-translation methods?
RQ2: How many words should learners read in order to improve their reading comprehension and reading fluency?
RQ3: How do the different instructional methodologies affect the amount of study time?
Hypothesis 1: Extensive reading plus an output activity better facilitates a group's reading comprehension and reading fluency than those of traditional intensive reading plus grammar-translation methods.
Hypothesis 2: The more the participants in the extensive reading plus an output activity group read, the better their comprehension and fluency levels will become.
Hypothesis 3: The total amount of study time for the extensive reading plus an output activity groups (D and E) is greater than the traditional intensive reading plus grammar-translation group.

METHODOLOGY

Participants

The participants of this research were 87 first-year female college students from the faculty of Economics (non-English majors). All the participants were native Japanese speakers enrolled in a General English course. The participants had studied English for six years in Junior and Senior high school and their English proficiency was mostly around levels A1 to B1 in the Common European Framework of Reference for Languages (CEFR). The preliminary questionnaire confirmed that over half of the participants had negative feelings about learning English, while 75 percent felt that they were not good at English. Although over 50 percent of the

participants answered that they liked to read Japanese books, only eight percent of them replied that they liked reading English. None of the participants had read English books, with the exception of school textbooks. The economics course offered two English-related subjects, one of which was compulsory General English, where this research was conducted, while the other was elective Business English. Business English was taken by 81 percent of the students. The participants had taken an average of 260 minutes of English classes a week in high school (before entering college), while studying for an average of 143 minutes a week outside classroom hours. They were divided into three groups in order to take General English classes based on the results of a placement test. Group A (31 students) was the most proficient, Group B (31 students) was the second most proficient, while the remaining students were placed in Group C (25 students). However, Groups B and C were reclassified on the basis of the amount of reading completed as D (over 50,000 words), and E (under 50,000 words). Table 1 shows the main points of the preliminary questionnaire results.

Table 1. Preliminary Questionnaire Results

		A ($n = 31$)		D ($n=20$)		E ($n=36$)	
		Freq	Pct	Freq	Pct	Freq	Pct
Do you like studying English?	5. Strongly agree	0	0.0	0	0.0	1	2.8
	4. Somewhat agree	11	35.5	3	15.0	3	8.3
	3. Neutral	9	29.0	3	15.0	13	36.1
	2. Somewhat disagree	8	25.8	12	60.0	13	36.1
	1. Strongly disagree	3	9.7	2	10.0	6	16.7
Are you good at English?	5. Strongly agree	0	0.0	0	0.0	0	0.0
	4. Somewhat agree	2	6.5	3	15.0	0	0.0
	3. Neutral	14	45.2	1	5.0	5	13.9
	2. Somewhat disagree	11	35.5	7	35.0	13	36.1
	1. Strongly disagree	4	12.9	9	45.0	18	50.0
Do you like reading Japanese books?	5. Strongly agree	5	16.1	3	15.0	4	11.1
	4. Somewhat agree	16	51.6	6	30.0	13	36.1
	3. Neutral	8	25.8	8	40.0	10	27.8
	2. Somewhat disagree	1	3.2	3	15.0	8	22.2
	1. Strongly disagree	1	3.2	0	0.0	1	2.8
Do you like reading English books?	5. Strongly agree	0	0.0	0	0.0	0	0.0
	4. Somewhat agree	3	9.7	2	10.0	2	5.6
	3. Neutral	13	41.9	6	30.0	4	11.1
	2. Somewhat disagree	10	32.3	6	30.0	12	33.3
	1. Strongly disagree	5	16.1	6	30.0	18	50.0
Do you take Business English?	Yes	28	90.3	16	80.0	25	69.4
	No	3	9.7	4	20.0	11	30.6

Material

Instructional method

For all three groups, classes consisted of 15 weeks per semester, with each lesson lasting for 90 minutes; the classes were taught separately by two different teachers. One teacher taught Group A, with the other teacher overseeing groups B and C. Group A was taught using an intensive reading method, which mainly focused on the Grammar Translation Method by using a textbook that included an approximately 300-word text with a vocabulary explanation followed by an approximately 140-word dialogue with composition drills focusing on particular grammatical points and formulaic expressions in each unit. The students were asked to read the text before the class and the instruction mostly focused on explanation of grammar and translation, which were considered to be typical GTM and did not focus on any particular reading skills such as skimming or scanning. The 300-word texts were narratives explaining famous cities and places such as Paris, New York, and San Francisco. The readability of one of the text was 68.8 (Flesh Kincaid Ease), and 8.8 (Flesh Kincaid Grade Level). The dialogues dealt with situations such as the front desk at a hotel, at the airport, and in a restaurant. The readability of the example dialogue was 82.8 (Flesh Kincaid Ease), and 4.8 (Flesh Kincaid Grade Level). 2000-word level vocabulary covered 93.47% for the former and 90.51% for the latter (VocabProfilers).

The class content and procedures were identical for Group B and C, which included in-class extensive reading for 30 minutes followed by a 10-minute book report session. The remainder of the class (50 minutes) included speaking activities using a textbook that included approximately 50-word texts introducing Japanese culture and life. Therefore, the content of the output activity had little to do with that of the Graded Readers. The text was a narrative whose readability was 97.3 (Flesh Kincaid Ease), and 1.3 (Flesh Kincaid Grade Level). A 2000-word level vocabulary covered 97.41% (VocabProfilers). The text was easy enough to not need grammar instructions. They read the text aloud three times and worked in pairs to translate an L1 version into an L2 version of the text and vice versa so that they were able to retell the text in the target language. The instruction did not focus on any particular reading skills such as skimming or scanning.

Questionnaires and Interviews

Two questionnaire surveys and informal interviews were conducted during the course. The preliminary questionnaire asked the participants about their learning, including their general attitude towards learning English and their previous English learning experiences, including their reading habits etc. This was issued during the first week of the first semester (April). The final questionnaire was conducted during the 12th week of the second semester (December), asking the class to share their experiences and their feelings regarding the course. This included asking Group A about their learning motives, while also asking Group B and C about their extensive reading experiences in the course, as well as similar items. In addition, questions regarding the number of hours the participants spent on English learning both during and outside class for all three groups before entering college and while in college were also asked.

Global Test of English Communication (GTEC) Academic Test

GTEC Academic was administered as a pretest during the first week of the first semester (April) and as a posttest in the 12th week of the second semester (December). This test, which was designed by Benesse Corporation and Berliz to target tertiary education, contains both Listening and Reading sections. It is a CBT online test and takes around 30 minutes to complete, with each section worth 250 points. The participants took both the reading and listening sections, however,

only the reading section was used in this research. The Reading section consists of three parts. Part A asks test takers to choose the appropriate word from four options to complete a sentence. Part B requires them to read an approximately 150-word passage and answer multiple choice questions which require skills such as summarizing and skimming. Part C consists of two approximately 350-word passages followed by three multiple choice questions, which include factual questions on details. Part B offers a wide range of text types such as emails, articles, and advertisements. Part C is expository. The readability of the example text for Part C was 42.1 (Flesh Kincaid Ease), and 13.2 (Flesh Kincaid Grade Level). A 2000-word level vocabulary covered 85.53% (VocabProfilers).

Reading Rate Test

Reading rate tests were conducted three times during the duration of the course. The first of these was conducted in the seventh week of the first term (May), with the second one occurring in the 15th week of the same term after Summer vacation (September). The final test was issued during the 12th week of the second term (December). Three different texts (A, B, and C) were used. These were taken from a training book by Notou and Teraguchi (2004) and modified by the author. The text types were narratives and the topics were familiar to the participants. The topics and the number of words were as follows: Text A: School life, 159 words; Text B: Family, 159 words; Text C: School life, 154 words. The readability of these texts and 2000-word level vocabulary coverage are shown in Table 2. In order to confirm the difficulty of the texts, reading rates were measured. A total of 30 English major students were recruited and randomly assigned to one of three groups, each of which consisted of 10 students. They were asked to read the three texts in different orders and answer each of the comprehension questions. One group started from text A, followed by text B and then text C. Another group adopted the order of B-C-A, while the final group used C-A-B. This was in order to minimize the effect of the order in which they read the texts. No one had a comprehension test score below three out of five points and the data was analyzed. Table 3 shows the average reading rate for each text. One-way Analysis of Variance (ANOVA) was conducted on the wpm for these texts, which confirmed that there was no significant difference among them. This meant they were considered to be of a homogeneous level in terms of difficulty, F (2, 89) = 1.905, p = .158 $n. s.$, r =.18. The comprehension questions consisted of five statements that were to be judged whether they were true or false according to the content of the text. The main purpose of using comprehension questions was that they made the participants read the text for meaning rather than just seeking speed while sacrificing comprehension. Therefore, the level of difficulty for the comprehension questions was not high. Still, setting the comprehension questions for each text is considered to be appropriate because reading speed without comprehension is worthless (Nuttall, 1996).

Table 2. Readability of Each Text

	Total Number of Words	Flesh Kincaid Ease	Flesh Kincaid Grade Level	2000-word Level Vocab Coverage
A (May)	159.0	73.6	6.0	93.76
B (Sep)	154.0	74.0	6.1	95.48
C (Dec)	143.0	74.3	6.3	96.55

Table 3. Mean Wpm for Each Text

	A		B		C	
	M	*SD*	*M*	*SD*	*M*	*SD*
N = 30	117.849	23.254	117.884	28.086	112.482	26.779

M-Reader System

The M-Reader system, an online management program for extensive reading courses first developed by Tomas Robb at Kyoto Sangyo University (Robb & Kano, 2013), was adopted only for groups B and C. The participants were required to take short, timed online quizzes on the book they read and, if they passed the instructor's preset percentage of correct answers, the number of words in the book was automatically added to the total amount of words each student read. 40,000 words was set as the minimum requirement for the academic year. However, the participants were also encouraged to read over the minimum requirement both during and outside the class period. The criterion for passing the quiz was set for six correct answers out of ten.

Graded Readers

Graded readers from various publishers such as Cambridge, Macmillan, Oxford, Pearson, and Penguin, were selected from the M-Reader book list. They were in the college library so that students could borrow them any time during its service hours. There were over 2,300 books, 65% of which were level zero through three in the M-Reader system. The readability of one book from level three was 91.8 (Flesh Kincaid Ease), and 2.0 (Flesh Kincaid Grade Level). A 2000-word level vocabulary covered 87.98% (VocabProfilers).This is considered to be crucial because the success of the course depended largely on the condition that students could access a large number of beginning level books (Day & Bamford, 2002). Besides this, the collection covered a variety of genres and topics to meet the needs of the students.

Procedure

The GTEC Academic test was administered in the first week of the first semester, followed by the preliminary questionnaire. A class orientation was also conducted, which included an explanation concerning what extensive reading was, how to do it, and how to use the M-Reader system. Normal classes continued from the second week until the 14th week (at the end of July). After the fourteenth week, the test was issued prior to Summer vacation. Group A was tested on the basis of the textbook, while Group B and C had a spoken test on the textbook. No instruction was provided during the Summer vacation, with the class then resuming from the fifteenth week (at the beginning of September). However, the class of the 15th week was spent reviewing the test results, meaning that no ordinary instruction, including extensive reading, was conducted. The second semester started the following week, with normal instruction continuing until the 11th week. The second GTEC Academic test was administered during the 12th week, along with the final questionnaire. An informal interview was performed during the 13th week to confirm the replies given on the questionnaire.

Data Collection Sessions

The participants were asked to gather at a computer room during the first week of the course, where each of them was allocated to a computer. Following the instruction, they took the GTEC Academic test, starting with the Listening section before proceeding to the Reading section. It was a fully computer-based test and everyone finished it within 30 minutes. The results were instantly sent to the test center, with the scores made available on the same day. Immediately after the test, the preliminary questionnaire was conducted. The participants were asked to answer via Google form. Reading rate was measured during the 7th week by asking the participants to read Text A while timing themselves using their smartphones. They then turned the paper over to answer five comprehension questions, judging whether each statement was true based on the content of Text A. The second reading rate measurement was conducted three months later (after Summer vacation) using the same procedure with Text B. The second GTEC Academic test was administered in the 12th week of the second semester, again using the same procedure. The final reading rate measurement was conducted following the test using Text C, with the same procedure applied once again. Immediately after the test, the final questionnaire was issued via Google form. A week after the final questionnaire, the informal interview took place during class time. The number of words each student had read was checked by the M-Reader system.

Data Analysis

Based on the data from the M-Reader, groups B and C were divided into Group D (Extensive reading High), which consisted of students who read more than 50,000 words, and Group E (Extensive reading Low), which consisted of those who read less than 50,000 words, for analysis. Group A (Intensive reading with GTM) was the control group. For the first source of data, the reading section of the two GTEC Academic tests (a pretest and a posttest) were compared using Analysis of Covariance (ANCOVA) to see if there were differences regarding their reading comprehension. As a post-hoc test, a Bonferroni test was also conducted. The second source of data consisted of the results from the three reading rate tests. The results (Wpm) of the first and the third measurements were compared using Analysis of Covariance (ANCOVA) to identify if there were differences regarding the reading rates of the different groups. In addition to this, Analysis of Variance (ANOVA) was conducted for the three measurements to test whether there were differences among the groups and to check the timing of each measurement. Ryan's method was also adopted for the post-hoc tests. The final source of data came from the results of the preliminary questionnaires, the final questionnaires, and an informal interview. The questionnaire consisted of a Likert scale, as well as free written responses.

RESULTS

An Analysis of Covariance (ANCOVA) was conducted on the scores of the Reading section of the GTEC Academic test for both the pre- and post-tests, with the instructional method (Intensive reading plus GTM vs. Extensive reading High plus output activity vs. Extensive reading Low plus output activity) used as a between-subjects independent variable for the groups with different proficiency levels. The results of pretest were covariates.

Table 4 shows the mean scores and standard deviations of the Reading section of the GTEC Academic tests for each condition, while Figure 1 highlights the mean scores. Table 5 shows the number of words read by groups D and E. The average number of quizzes D and E took (the number of books they read) and the number of quizzes they failed, whose number of words were not counted, were as follows: Group D took 113.00 quizzes and failed 8.85 (92.17% pass), and Group E took 82.62 quizzes and failed 5.49 (93.36% pass) under the criteria of six correct answers out of ten.

Table 4. Descriptive Statistics for Each Group on the Pre- and Post-test

	Pretest		Posttest	
	M	*SD*	*M*	*SD*
A (IR: *n* = 31)	92.13	21.37	88.35	19.14
D (ER-H: *n* = 20)	63.20	19.66	58.00	25.08
E (ER-L : *n* = 36)	60.61	22.52	54.44	19.71
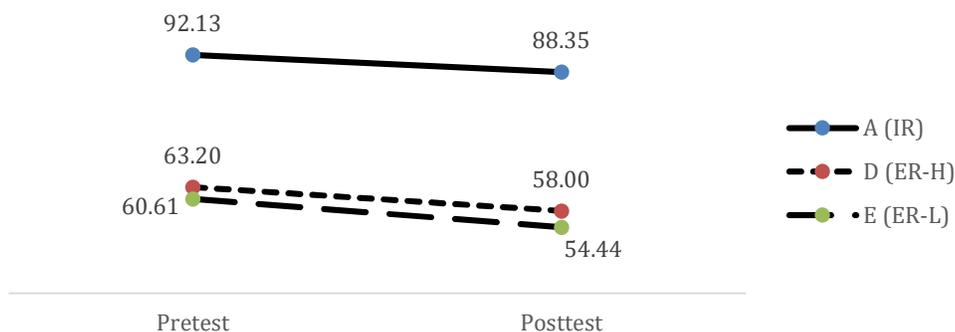
Figure 1. GTEC Academic Test (Reading Section)

Table 5. The Number of Words Group D and E Read

	M	*SD*	Max	Min
D (ER-H: *n* = 20)	74,029.30	27,326.12	157,742.00	50,068.00
E (ER-L : *n* = 36)	35,379.28	9,087.97	49,839.00	14,557.00

The analysis revealed significant differences among the groups in the pretest with large effect size, $F (1, 86) = 71.322$, $p < .001$, $\eta2 = .462$. It also revealed significant differences among the groups in the posttest, meaning there was a significant impact on the different instructional methods with small effect sizes, $F (2, 83) = 4.476$, $p = .014$, $\eta2 = .097$ (Table 6). A Bonferroni Post Hoc Test was conducted, showing that there was also a significant difference between Group A (IR) and Group E (ER-L) with large effect sizes, $t (65) = 13.117$, $p = .014$, $r = .85$.

Table 6. Tests of Between-Subjects Effects

Dependent Variable: Source	Type lll Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncentral Parameter	Observed Power
Corrected Model	38304.490	3	12768.163	53.939	0.000	0.661	161.818	1.000
Intercept	2593.097	1	2593.097	10.955	0.001	0.117	10.955	0.905
Pretest	16882.821	1	16882.821	71.322	0.000	0.462	71.322	1.000
Group	2119.014	2	1059.507	4.476	0.014	0.097	8.952	0.752
Error	19647.165	83	236.713					
Total	452525.000	87						
Corrected Total	57951.655	86						

R Squared = .661 (Adjusted R Squared = .649)
Calculated with Alpha = .05

      The mean and standard deviation of the reading rates (wpm) for the respective groups across three different periods is shown in Table 7, while Figure 2 highlights the mean scores. An Analysis of Covariance (ANCOVA) was conducted on the first (May) and last (Dec) measurements. The results showed there was no significant difference among the groups concerning score changes in mean reading rate between May and December, $F(2, 77) = 1.883$, $p = .159$, n.s., $\eta2 = .047$, although the reading rates in May produced significant differences, $F(1,77) = 11.076$, $p < .001$, $\eta2 = .126$. Furthermore, an Analysis of Variance (ANOVA) was conducted on reading rates for the groups between the subject variable and the time (May, Sep, and Dec) within the subject variable. The results also highlighted a significant difference among the groups with large effect sizes, $F(2, 78) = 14.577$, $p < .001$, $\eta2 = .61$. Ryan's method was adopted as a post-hoc test, revealing that there was a significant difference between groups A and D with medium effect sizes, $t(78)$ 3.169, $p < .01$, $r = .34$, as well as between groups A and E with large effect sizes, $t(78) = 5.863$, $p < .001$, $r = .55$. The results also revealed significant differences regarding time with large effect sizes, $F(2, 156) = 36.611$, $p < .001$, $\eta2 = .47$. Ryan's method as a post-hoc test indicated that there was a significant difference between May and December with large effect sizes, $t(156) = 8.293$, $p < .001$, $r = .55$, and between September and December with medium effect sizes, $t(156) = 6.719$, $p < .001$, $r = .47$.

Table 7. Descriptive Statistics for Each Group on Reading Rates (Wpm)

|	May		Sep		Dec	
	M	SD	M	SD	M	SD
A (IR: n = 28)	98.072	24.268	99.832	19.576	117.411	22.981
D (ER-H: n = 19)	79.621	24.101	85.574	23.821	100.566	23.23
E (ER-L : n = 34)	68.817	17.451	73.567	20.269	94.205	22.754
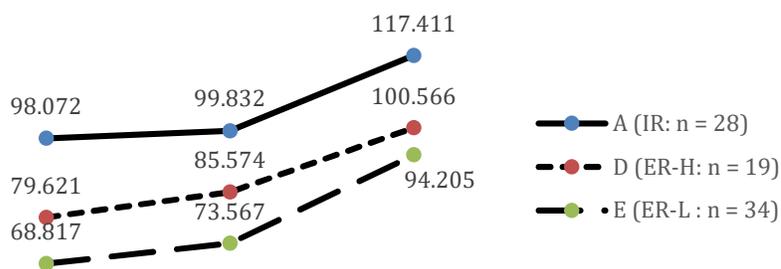
Figure 2. Reading Rates (Wpm)

Tables 8 and 9 show the results of the final questionnaire, which concerned both how groups D and E felt about their extensive reading experiences and how each group felt about English studies and the course.

Table 8. The Final Questionnaire on Extensive Reading 1

		D (*n* = 20)		E (*n* = 36)	
		Freq	Pct	Freq	Pct
I read books extensively.	5. Strongly agree	9	45.0	2	5.6
	4. Somewhat agree	10	50.0	9	25.0
	3. Neutral	1	5.0	12	33.3
	2. Somewhat disagree	0	0.0	10	27.8
	1. Strongly disagree	0	0.0	3	8.3
Extensive reading was interesting.	5. Strongly agree	4	20.0	2	5.6
	4. Somewhat agree	8	40.0	10	27.8
	3. Neutral	5	25.0	12	33.3
	2. Somewhat disagree	2	10.0	9	25.0
	1. Strongly disagree	1	5.0	3	8.3
I felt fulfilled after I finished reading books.	5. Strongly agree	11	55.0	16	44.4
	4. Somewhat agree	7	35.0	9	25.0
	3. Neutral	0	0.0	8	22.2
	2. Somewhat disagree	2	10.0	3	8.3
	1. Strongly disagree	0	0.0	0	0.0
I can read faster through extensive reading.	5. Strongly agree	5	25.0	4	11.1
	4. Somewhat agree	9	45.0	16	44.4
	3. Neutral	4	20.0	11	30.6
	2. Somewhat disagree	2	10.0	5	13.9
	1. Strongly disagree	0	0.0	0	0.0
I read books because it was a requirement for the class.	5. Strongly agree	7	35.0	15	41.7
	4. Somewhat agree	11	55.0	18	50.0
	3. Neutral	1	5.0	1	2.8

	2. Somewhat disagree	0	0.0	2	5.6
	1. Strongly disagree	1	5.0	0	0.0
The M-Reader system was easy to use.	5. Strongly agree	7	35.0	12	33.3
	4. Somewhat agree	11	55.0	14	38.9
	3. Neutral	2	10.0	6	16.7
	2. Somewhat disagree	0	0.0	2	5.6
	1. Strongly disagree	0	0.0	2	5.6

Table 9. The Final Questionnaire on Extensive Reading 2

		A ($n = 31$)		D ($n = 20$)		E ($n = 36$)	
		Freq	Pct	Freq	Pct	Freq	Pct
It is interesting to know about people from English-speaking countries and their ways of life.	5. Strongly agree	6	19.4	8	40.0	4	11.1
	4. Somewhat agree	12	38.7	6	30.0	15	41.7
	3. Neutral	8	25.8	2	10.0	13	36.1
	2. Somewhat disagree	4	12.9	4	20.0	3	8.3
	1. Strongly disagree	1	3.2	0	0.0	1	2.8
I feel happy when I keep studying and become able to hear and understand the words and phrases that I could not previously recognize.	5. Strongly agree	2	6.5	3	15.0	6	16.7
	4. Somewhat agree	22	71.0	10	50.0	17	47.2
	3. Neutral	4	12.9	4	20.0	10	27.8
	2. Somewhat disagree	1	3.2	3	15.0	2	5.6
	1. Strongly disagree	2	6.5	0	0.0	1	2.8

Table 10 shows the weekly study time for the participants during a typical week in both the third year of high school and the first year of junior college, while Figure 3 shows the differences regarding the mean length of time.

Table 10. The Amount of Time Spent on English Studies per Week (Minute)

	High School						Junior College					
	School		Home		Total		School		Home		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
A (IR: $n = 31$)	275.65	101.36	202.26	194.88	477.90	244.36	171.29	27.05	58.39	61.65	229.68	62.95
D (ER-H: $n = 20$)	232.00	71.95	97.50	75.32	329.50	118.55	171.00	27.70	95.00	50.21	266.00	55.67
E (ER-L : $n = 36$)	256.11	94.93	110.83	213.68	366.94	245.54	147.50	43.84	81.11	71.14	228.61	85.53

477.90

366.94

329.50

266.00

229.68

228.61

● A (IR)
● D (ER-H)
● E (ER-L)
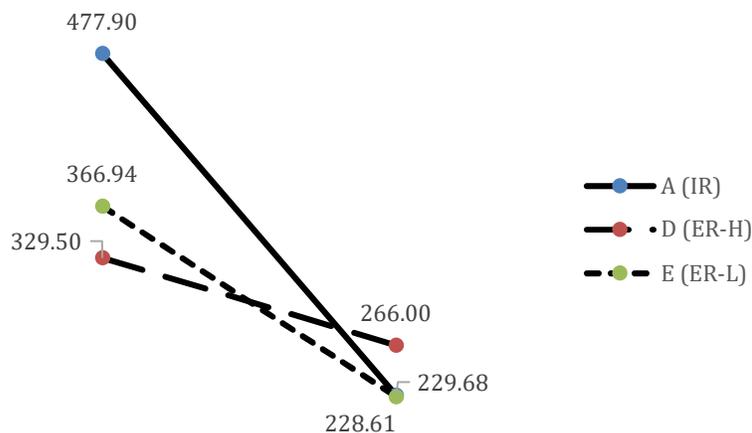
Figure 3. Weekly study time for each group

A 2×3 Analysis of Variance (ANOVA) was conducted on their study time for a typical week consisting of two periods: one when they were in their third year of high school and the other when they were in their first year of junior college. The results suggested there was no significant difference among groups with medium effect sizes, $F (2, 84) = 1.74$, $p = .182$, *n.s.* However, there were significant differences between the two periods (high school and junior college) with large effect sizes, $F (2, 84) = 41.820$, $p < .001$, $\eta2 = .498$, while there was also significant interaction between groups and periods with medium effect sizes, $F (2, 84) = 5.347$, $p < .01$, $\eta2 = .127$. Therefore, a simple main effect test was conducted, which revealed a significant difference between the groups regarding high school study time with a large effect sizes, $F (1, 84) = 5.928$, $p < .005$, $r = .87$. Ryan's method was conducted as a post-hoc test, showing a significant difference between groups A and D with medium effect sizes, $t (49) = 3.126$, $p < .005$, $r =.41$, and between groups A and E with medium effect sizes, $t (65) = 2.736$, $p < .01$, $r = .32$. However, no significant difference was found between groups D and E with small effect sizes, $t (54) = 0.811$, $p = .418$ *n.s.*, $r = .11$. This means that group A studied for significantly longer than the other groups. A simple main effect test also revealed a significant difference between high school and junior college for group A with large effect sizes, $F (1, 84) = 38.165$, $p < .001$, $r = .987$, and for group E with large effect sizes, $F (1, 84) = 11.853$, $p < .001$, $r = .960$. However, no significant difference was found for group D with small effect sizes, $F (1, 84) = 2.498$, $p = .118$, *n.s.*, $r = .17$. The time spent studying in junior college dropped significantly only for groups A and E, though the study time in junior college for group D dropped similarly compared to the other groups. The results of the informal interviews are discussed in the Discussion section.

DISCUSSION

RQ1 asked whether extensive reading plus output activity better facilitates reading comprehension and reading fluency than traditional intensive reading plus grammar-translation methods, while RQ2 aimed to investigate how many words learners should read in order to improve their reading comprehension and reading fluency. Regarding these RQs, Hypothesis 1 presumes that extensive reading plus output activity better facilitates learners' reading comprehension and reading fluency. In terms of reading comprehension, all three groups' scores

were reduced on the posttest. A significant difference was found only between groups A (IR) and E (ER-L), which means that extensive reading plus output activity was as effective as conventional instruction if the participants read over 50,000 words (approximately 70,000 average). This result was in line with previous research, although the scores were more significantly reduced (Fujita & Noro, 2009; Nakanishi & Ueda, 2011; Robb & Susser, 1989). One possible reason for the decline in scores could be that the total English study time was greatly reduced after students enter junior college compared to their high school days, and this was especially true of groups A (IR) and E (ER-L). Another possible reason is that the test was designed to assess academic ability and was thus too difficult for them, which was also reported in previous research (Nakanishi & Ueda, 2011). The Flesh Kincaid Ease scores indicated that all the text but the GTEC Academic test, namely the text books for Group A, D, and E, Graded readers of, for example, level 3 books, and reading rate tests, were between fifth grade to eighth grade level, while the GTEC Academic test was college level. A 2000-word level vocabulary covered 85.43 for the GTEC Academic test, while 87.98 % to 97.41 % for the rest of the texts. Another possible reason is that it takes time to improve reading ability compared to reading fluency (Yamashita, 2008). All three groups significantly improved, though no significant difference was reported between May and September. One possible reason for this was that the test in September was conducted right after they resumed class after one month of Summer vacation, which means most of the students did not read or study English for a long time. Therefore, Hypothesis 1 was not supported. There was no significant difference among the groups regarding how they improved fluency, though there were significant differences in overall reading speeds between group A (IR) and the other groups (D:ER-H, E:ER-L). That there was no significant difference between groups D and E could mean that quantity of reading has little to do with the incremental improvement of fluency, and this position is also supported by prior research (Huffman, 2014; Nakanishi & Ueda, 2011).

Hypothesis 2, stating that the more the participants read the better their comprehension and fluency would become, was also not supported. In addition to the previously mentioned Summer vacation, during May, Group A had as few as five classes, while groups D and E had already read an average of 11,713 words. This means they could have already improved their reading speed.

RQ3 questioned if a difference in instructional methodologies affects the amount of study time. Hypothesis 3, stating that the total amount of study time for groups D and E will be greater than Group A, was also not supported. All of the groups reported drops in study time compared to their high schools, however, only group D (ER-H) reported no significant difference between these study periods. As Berwick and Ross (1989) point out, the incentive to study English peaks in high school in Japan. Therefore, an appropriate intervention to maintain their incentive to read English should be necessary. Although Groups D and E reported similar traits in the results of the questionnaire, Group D had a tendency to feel more fulfilled and to perceive extensive reading instruction positively. Partly because Group D was more proficient and motivated, and partly because of Japanese culture, where people think they could have been better no matter what. Especially, there were more students who barely read the required number of words. Also, the most proficient group (A) had more positive feelings regarding both the English-speaking community and English learning in general. Multiple students pointed out in their interviews that studying English, including extensive reading, was hard, but they felt this was a fulfilling experience. They also added that, if there were no such activity, they would not have visited the library so often. Therefore, the instruction could have a good influence on non-English major students who do not enjoy studying English.

Pedagogical Imprecations

In order to improve reading comprehension, the number of words the learners read should be very important. However, in addition to this, the total study time, namely the amount of time the learners are exposed to English, should also be crucial. The questionnaire indicated the majority of students felt fulfilled after reading graded readers. Therefore, extensive reading has potential to be an idealistic methodology because it could make their study time less unpleasant and more enjoyable. Another important point is that teachers should not only encourage students but also use extensive reading a class requirement, and set a certain number of words to read because, as the questionnaire indicated, the learners might not read extensively unless it was required.

Limitations

This research was conducted as part of normal class instruction including periods in which the participants did not read for over a week or a month due to holidays. Actually, the second measurement of the participants' reading rate was measured right after one month of Summer vacation. Also, three intact class units were used for research. Therefore, there were significant differences in the participants' English proficiency including reading comprehension ability and reading rate among the groups. In addition, due to the setting, it was quite difficult to adopt a cross-treatment design.

Although a standardized test was preferable, the text type and their readability differed greatly between the instructional materials and the GTEC Academic test. Furthermore, the first reading rate test should have been conducted before the onset of the course in order to measure initial reading speed.

CONCLUSION AND FUTURE RESEARCH

This study aimed to investigate whether the instructional method of extensive reading plus output activity better facilitates the participants' reading comprehension and reading fluency compared to intensive reading plus traditional grammar-translation methods. It also tried to offer a possible instructional model that includes extensive reading for Japanese EFL classrooms. The results suggest that total study time plays an important role in improving reading comprehension, and that learners should read at least 50,000 words in order to be as effective as they could be through conventional teaching methodologies. Furthermore, reading fluency can also be improved, regardless of the form of instruction or the total amount of words read. Future research should shed light on other areas related to English learning, such as vocabulary size and learning motivation related to extensive reading.

ACKNOWLEDGEMENT

REFERENCES

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language, 26*, 29-48.

Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning, 62*, 665-703.

Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix, 1*. http://www.readingmatrix.com/articles/bell/index.html

Berwick, R., & Ross, S. (1989). Motivation after matriculation: Are Japanese learners of English still alive after exam hell? *JALT Journal, 11*, 193-210.

Breznits, Z. (1988). Reading performance of first graders: The effects of pictorial distractors. *The Journal of Educational Research, 82*, 47-52.

Brown, P., & Hirst, S. B. (1983). Writing reading courses: The interrelationship of theory and practice. In Brumfit, C.J. (Ed.) (1983). *Language teaching projects for the third world*. ELT documents 116. British Council English Teaching Information Centre.

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

Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language, 14*, 136-141.

Fujita, K., & Noro, T. (2009). The effects of 10-minute extensive reading on the reading speed, comprehension and motivation of Japanese high school EFL learners. *Annual Review of English Language Education, 20*, 21-30.

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

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

Grabe, W., & Stoller, F. L. (2002). *Teaching and Researching Reading*. Edinburgh: Pearson Education Limited.

Hagley, E. (2017). Extensive graded reading with engineering students: Effects and outcomes. *Reading in a Foreign Language, 29*, 203-217.

Hayashi, K., (1999). Reading strategies and extensive reading in EFL classes. *RELC Journal, 30*, 114-132.

Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language, 26*, 17-33.

Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language, 20*, 70-91.

Jeon, E., & Day, R. R. (2015). The effectiveness of core ER principles. *Reading in a Foreign Language, 27*, 246-265.

Jeon, E., & Day, R. R. (2016). The effectiveness of ER on reading proficiency: A meta-analysis. *Reading in a Foreign Language, 28*, 246-265.

Kanatani, K., Takayama, Y., Usukura, M., & Ota, E. (2011). *Koko eigo jyugyo wo kaeru* [Let's change English classes at high school]. Tokyo: ALC Press Inc.

Lai, F. K. (1993). The effect of a summer reading course on reading and writing skills. *System, 21*, 87-100.

Macalister, J. (2010). Investigating teacher attitudes to extensive reading practices in higher education: Why isn't everyone doing it? *RELC Journal, 41*, 59-75.

Maley, A. (2005). Review: Extensive reading activities for the second language classroom. *ELT Journal, 59*, 354-355.

Matsui, T., & Noro, T. (2010). The effects of 10-minute sustained silent reading on junior high school EFL learners' reading fluency and motivation. *Annual Review of English Language Education, 21*, 71-80.

Ministry of Education, Culture, Sports, Science and Technology (2017) Announcement on FY2016 School basic survey (Confirmed value). http://www.mext.go.jp/en/news/topics/detail/__icsFiles/afieldfile/2017/10/13/1397166_001_1.pdf (accessed 2019. June. 21.)

Nakanishi, T., & Ueda, A. (2011). Extensive reading and the effect of shadowing. *Reading in a Foreign Language, 23*, 1-16.

Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. New York, NY: Routledge.

Notou, S., & Teraguchi, H. (2004). *Sonic reading stage zero starter*. Tokyo: Kiriharashoten.

Nuttall, C. (1996). *Teaching reading skills in a foreign language* (2nd Ed.). Oxford: Heinemann English Language Teaching.

Park, J. (2015). Integrating reading and writing through extensive reading. *ELT Journal, 70*, 287-295. DOI: 10.1093/elt/ccv049.

Ro, E. (2014). Pleasure reading behavior and attitude of non-academic ESL students: A replication study. *Reading in a Foreign Language, 26*, 49-72.

Robb, T., & Kano, M. (2013). Effective extensive reading outside the classroom: A large-scale experiment. *Reading in a Foreign Language, 25*, 234-247.

Robb, T. N., & Susser, B. (1989). Extensive reading vs skills building in an EFL context. *Reading in a Foreign Language, 5*, 239-251.

Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System, 36*, 661-672.

Akira Iwata *is an Associate Professor at Hokkaido Musashi Women's Junior College. He previously taught at several public high schools for 21 years. He holds a Master of Education in TESOL from the University of Wollongong, Australia. His research interests include EFL classroom practices, particularly EFL pedagogy.*
Email: iwata@hmjc.ac.jp