Quantitative Assessment of the Extensive Reading Program in the Global English Department, with a Focus on the Development of General Reading Ability and Linguistic Proficiency

Tadashi NORO, Gregory Lewis ROHE, Sayuri OKADA, Shinobu HAMADA

1. Introduction

Almost ten years have passed since the Extensive Reading Program started at the foundation of the Department of Global English. A lot of students have mentioned extensive reading as one of the most interesting and beneficial English language courses. However, some English teachers seem to wonder if extensive reading is really beneficial. Grabe (2009, p. 312) explains about the unpopularity of extensive reading, pointing out that "fluent reading is often not really the goal for a reading class or a reading curriculum; rather, the goal is the development of language skills, vocabulary, and grammar, translation, or study skills." Now it is high time to assess the program and clarify whether extensive reading instruction is an effective method for increasing learners' general reading ability and overall English language proficiency.

1.1 The definition of extensive reading

Extensive reading (ER) has received growing interest over the past two decades as a promising method for improving learners' reading fluency and language proficiency. What is ER? EPER (1992, p. 9) defines ER as follows: "By extensive reading, we mean above everything else reading in quantity. For this there are three conditions: a good reading speed and materials that both engage the interest of the student and are written at a level that matches the student's proficiency." Richards and Schmidt (2002, pp. 193-194) add some ideas: "Extensive reading means reading in quantity and in order to gain a general understanding of what is read. It is intended to develop good reading habits, to build up knowledge of vocabulary and structure, and to encourage a liking for reading." The key characteristics drawn from the two definitions are 1) reading in quantity, 2) reading at a good reading speed, 3) reading materials that are interesting and appropriate for students' proficiency, 4) reading materials that promote a positive attitude and motivation for reading, 5) to build up linguistic knowledge.

Grabe (2009, pp. 311-312) claims that "the ability to read extended texts for long periods of time is a hallmark of fluent reading... the only way to ensure the latter goal (fluent reading)....not by fluency practice but by practice in extensive reading." Grabe (1991) was not keen on the term *pleasure reading*, which Mikulecky used, mainly because the term is too specific and partly because the term has a frivolous tone for students, parents, and administrators. Later, Krashen (1993) used another term, *free voluntary reading* in his book *The Power of Reading*. Thus, it is important to note that a primary characteristic of ER can be *willingness to read a lot of easy extended texts for a long time.*

1.2 The characteristics of an extensive reading approach

Although various ways of implementing ER are suggested, Day and Bamford (1998, pp. 7-8) posited ten principles that were found in successful extensive reading programs, which are a helpful guideline for teachers to implement in an ER program.

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

In addition to these principles, three principles or tips were suggested by Sakai (2004):

1. Don't look up words in the dictionary.
2. Skip over parts you don't understand.
3. If you aren't enjoying one book, toss it aside and get another.

Sakai also suggests that teachers should have learners avoid consulting dictionaries and train them to skip unknown words, which encourages them to infer the meaning of words in context.

However, these are thought to be guidelines rather than principles, and we don't need to adhere to them with a single-minded rigidity. If students come across an unknown word several times, teachers may encourage them to look up the word in the dictionary after inferring its meaning from the context. Recently the third principle was slightly changed into the following guideline: If students don't progress smoothly while reading a book, stop it and find another, probably because the book might be difficult, uninteresting or boring.

1.3 Theoretical background

The theoretical framework for supporting ER was provided by *the input hypothesis* (Krashen, 1985, pp. 2-3), which posits that the process of exposure to *comprehensible input* is the necessary and sufficient cause of second language acquisition, and by *the reading hypothesis*, which is summarized in this way: " In-school free reading studies and out-of-school self reported free voluntary reading studies show that more reading results in better reading comprehension, writing style, vocabulary, spelling, and grammatical development (Krashen, 1993, p. 12)." In L2 settings, research has reported the beneficial effects of ER on various aspects of L2 ability, such as reading comprehension, reading speed, vocabulary, grammar, reading and writing, spelling and general L2 proficiency. However, it is uncertain whether all aspects of English proficiency can be influenced in EFL settings in Japan.

Comprehensible input is input language which contains linguistic items that are slightly beyond the learner's current linguistic competence (Richards and Schmidt, 2002, p. 99). The input hypothesis also says that input must contain $i + 1$ to be useful for language acquisition (Krashen, 1985, p. 21). "*i*" in $i + 1$ is an abbreviation of interlanguage, which means the student's current level of proficiency. In contrast to the comprehensible input hypothesis, Day and Bamford (1998, pp. 16-17) reformulated this hypothesis, and suggested that learners should read $i - 1$ materials for *automaticity training* (Samuels, 1994, p. 834), whose "goal is developing a large sight vocabulary rather than the learning of new linguistic elements (p. 16)." Nuttall (2005, p. 175) tried to show the acceptable proportion of new words in a text, and recommended the proportion of new words in a text in this way: "... For intensive reading ...it may be acceptable to accept quite a lot of new words.... 2 or 3 per cent is a lot in my view.... books for extensive reading should have a smaller proportion of new words - 1 per cent perhaps." Regardless of whether a text is at an $i + 1$ level or $i - 1$ level, students will read a great deal if they do so with tolerable ease.

Why is it necessary for Japanese students to be exposed to a great quantity of print in the target language? Figure 1 compares the number of tokens appearing in middle school English textbooks in Korea, Taiwan, China, and Japan. Figure 2 also shows the number of tokens used in high

school English textbooks in Korea, Taiwan, China, and Japan. These two figures prove that Japanese students are exposed to far fewer English words than those in Taiwan, China, and Korea (especially in the case of middle school). Table 1 indicates the proportion of basic vocabulary used in high school English textbooks in Korea, Japan, China and Taiwan. The table reveals a marked decrease in the proportion of exposure to 1000 word level (77.98%) and 2000 word level (45.54%) in Japan, compared with those in Korea (89.81% & 63.73% respectively), China (84.16% & 52.72% respectively) and Taiwan (96.71% & 87.02% respectively). Considering the English teaching situation in Japan, it is to be hoped that ER is widely adopted in the curriculum to increase the exposure to basic vocabulary.
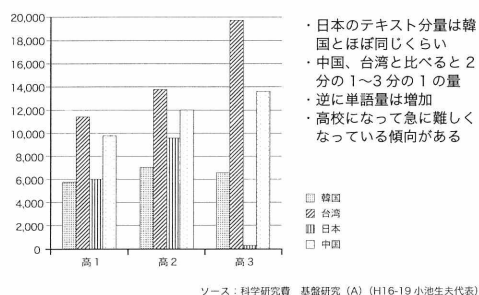

Figure 1. The comparison of tokens in middle school text-books among Korea, Taiwan, China and Japan. (Tono, 2015, p. 40)
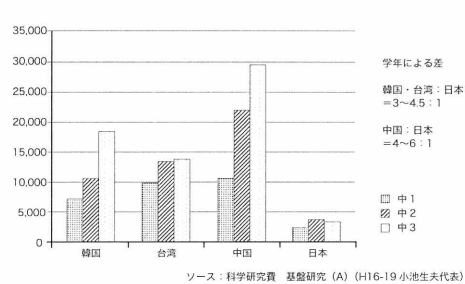

Figure 2. The comparison of tokens in high school text-books among Korea, Taiwan, China and Japan. (Tono, 2015, p. 42)

Table 1
The Proportion of Exposure to Basic Vocabulary (Tono, 2015, p. 43)

BNC 頻度ランク	使用言語	韓国	日本	中国	台湾
1,000	972	89.81%	77.98%	84.16%	96.71%
2,000	863	63.73%	45.54%	52.72%	87.02%
3,000	656	43.14%	24.70%	32.01%	76.98%
4,000	454	37.67%	16.96%	23.57%	68.94%
5,000	342	35.96%	19.30%	14.91%	69.94%
6,000	250	30.40%	14.40%	18.40%	66.80%
7,000	150	33.33%	8.00%	16.00%	63.33%
8,000	129	28.68%	10.08%	9.30%	69.77%
9,000	93	30.11%	10.75%	10.75%	70.97%
10,000	70	27.14%	12.86%	15.71%	62.86%

1.4 Previous studies

Previous studies on the effects of extensive reading were administered focusing on the effects on different aspects of language abilities/skills, such as reading comprehension ability, vocabulary, writing, grammar and spelling. The present research concerns general reading ability and overall English language proficiency developed by the extensive reading practice in Japanese EFL settings. Therefore, this literature review is confined to the studies on the effects on reading

comprehension and English language ability.

Hayashi (1999) investigated the effects of extensive reading on the four sub-sections of the TOEFL test (reading comprehension, vocabulary, structure and writing, and listening), and the use of reading strategies. The participants were 100 Japanese university students, who received extensive reading instruction for approximately 10 months. The reading comprehension section displayed larger improvement than the other sections. The study did not discuss why the TOEFL test was used to measure reading comprehension ability developed through extensive reading instruction.

Tanaka & Stapleton (2007) explored the effects of a quasi-extensive reading program on reading comprehension and they reading speed. The participants in the treatment group were 96 Japanese high school students, and they read easy teacher-made materials for 5 to 10 minutes at the beginning of every English class for approximately 5 months. Some of these students read graded readers outside of class. The treatment group, especially those who read graded readers, scored significantly higher in reading speed and comprehension than the control group.

Iwahori (2008) investigated the effects of extensive reading on reading rate and language proficiency. The participants were 33 high school students in Japan, and were provided with graded readers as homework for 7 weeks. Reading rate and general language proficiency (using C-test) improved significantly. However, it is a debatable point whether the researcher scored language ability instead of reading comprehension accuracy, in that fluent reading is the amalgamation of reading speed and comprehension accuracy.

Yamashita (2008) explored the effects of extensive reading on general reading ability and lower-level linguistic ability. The participants were 38 Japanese university students who received extensive reading instruction for one semester. The improvement of general reading ability was statistically significant, but not of linguistic ability. It is worth noting her conclusion that "the benefit of extensive reading is more likely to be manifest in general reading ability....and that progress in micro-level linguistic ability such as vocabulary, spelling and morphosyntax may follow later. (p. 669)."

1.5 Research questions

The purpose of the present study is to investigate the effects of ER practice for approximately 10 months on two aspects of L2 English ability. One is to examine the improvement of extensive reading ability, or general reading comprehension ability, and the other is to explore the development of overall English language proficiency or lower-level linguistic ability. To which aspect of L2 ability does ER practice have a more beneficial effect? Taking these purposes into account, the following research questions (RQs) were addressed in this study.

1. Does ER instruction for 10 months improve overall English language proficiency?

2. Does ER instruction for 10 months develop extensive reading ability?

3. On which ability does ER instruction for 10 months have more beneficial effects, overall English proficiency or general reading ability?

2. Method

2.1 Participants

The participants in this program were 120 first year students in the Department of Global English. However, those who missed either or both of the tests conducted in April as a pretest, or in December as a post-test were excluded from this study. As a result, 98 students took EPER placement tests (cloze tests), and 98 students (some were different from those involved in cloze tests) had extensive reading tests. Extensive Reading is a required course for the students in the Department of Global English.

2.2 Materials

Two kinds of tests developed by the Edinburgh Project on Extensive Reading (EPER) were used in this research as pretests and post-tests: 1) *placement/progress tests (PPT)* and 2) *extensive reading tests (ERT)*.

2.2.1 *Placement/progress tests (PPTs)*

PPTs are cloze tests, each of which is composed of more than 10 graded passages of increasing difficulty. There are three parallel versions available (Version A, Version B and Version E, though Version A seems to be easiest), of which Version A was used in the present study. There are 141 blanks in Version A. Each correct answer is scored 1. The reliability (Cronbach's alpha of Version A) was 0.78. The scores of cloze tests are used to indicate learners' reading ability level, to determine their starting levels of graded readers, which EPER classifies into eight levels of difficulty (X, A, B, C, D, E, F, G, from the advanced to the elementary) and to monitor their progress.

PPTs are not strictly fixed ratio cloze tests (namely, every nth word is not deleted) but semi-fixed ratio tests. Yamashita (2008, p. 665) analyzed this PPT and pointed out interesting features of this test as follows:

The first item is usually around 10th word at the beginning of each passage. After that, approximately every 5th to 8th word is deleted. An important characteristic of PPTs, ..., is

the relatively short length of passages.

From these characteristics it can safely be said that EPER intended to distinguish a wide variety of testees with different English comprehension ability.

PPTs are scored according to the Answer Key provided by EPER, in which only one answer is given to
many blanks but some blanks can accommodate alternative lexical or grammatical variations. The words which are incorrectly spelt are accepted if they are semantically acceptable. It can be said that PPTs are basically scored by an exact-word method, but try to combine an acceptable-word method within this limited range.

The same test (Version A) was administered as a pretest and as a post-test in this study. Nakanishi (2015) points out that this is "a design feature that potentially creates test-retest effects, particularly given that the treatment period in 63% of the studies was fewer than 6 months." Unlike reading comprehension tests, cloze tests seem to have fewer test-retest effects, because the tests consist of short passages, and lay more emphasis on lower-level linguistic competence, so that the students seemed not to remember the content of the test after a 9-month treatment.

The purpose of using the cloze test-type PPT A Level is to "measure reading comprehension as well as overall language proficiency (Richards & Schmidt, 2002, p. 78)," although there has been a controversy as to whether cloze tests measure reading comprehension ability or lower-level language ability (including ability to use grammatical knowledge and lexical knowledge in the context). According to Oller Jr. & Conrad (1971, p. 190), the correlation coefficient between cloze test and reading comprehension was 0.80, between cloze test and vocabulary 0.59, and between cloze test and grammar 0.58, though this research is rather outdated, and conducted with ESL learners, not EFL learners. On the other hand, Urquhart and Weir (1998) referred to a rational cloze test (a selective gap-filling test in their term) as a method of measuring overall English language proficiency. Regardless of discrepancies over the assessment of cloze tests, it can safely be said that cloze tests can measure overall English language proficiency, including reading ability and lower-level linguistic abilities.

2.2.2 *Extensive reading tests (ERTs)*

ERTs are original or unique in that they aim to assess *extensive reading ability*. In many cases of extensive reading research, to assess the development of reading ability after an extensive reading program, TOEFL, TOEIC or STEP styles of reading tests were constructed mostly based on short expository texts. In the light of the definition of extensive reading, these reading tests

don't reflect reading in quantity, a good reading speed, or reading materials that arouse the interest of readers and match the student's proficiency. Since ERTs aim at assessing extensive reading ability or general reading ability, the texts for ERTs are taken from English graded readers and are relatively long (Level D Version 1 - approximately 1,350 words with 20 questions to answer in 30 minutes; Level F Version 2 - approximately 950 words with 22 questions). Most of the questions are global comprehension questions. There are eight levels in ERTs (X, A, B, C, D, E, F, G, from the advanced to the elementary, as mentioned above). Various types of comprehension questions are included: yes/no questions, multiple-choice questions, questions demanding short answers in English, gap-filling or gap-summary questions.

2.3 Procedure

The extensive reading program was implemented by four instructors (Noro for A class, Rohe for B class, Okada for C class, and Hamada for D class) for 15 weeks in the spring semester and 15 weeks in the fall semester. The class met once a week for 90 minutes. The participants were expected to borrow from the University library whatever graded readers they wanted to read, to read them, and to submit a book report for each book. In the book report, they were required to fill in the name of the book read, the number of words and pages they read, comment on the book in terms of interest and ease/difficulty, and to write a report either in Japanese or in English, choosing one of the themes, such as a character, to recommend or not, a brief summary, favorite expressions, cultural differences, and sad or surprising events. They were encouraged to listen to a CD if it was attached to a graded reader (This practice is called *extensive listening*.)

Students read mainly outside of the class, and are also engaged in *sustained silent reading* in class for 30 minutes or so. It is important that students who have no experience in reading long texts can cultivate reading habits and get accustomed to extensive reading in class. Sustained silent reading in class can also give teachers a chance to monitor students reading. What level or types of books are they reading? Who is a fast reader? Who seems not to be concentrating? This kind of information is invaluable in helping teachers guide students in overcoming whatever barriers they face.

In-class activities can promote extensive reading and help students begin to see reading as a valuable and pleasurable activity and as a resource for developing other skills. The activities basically include 1) vocabulary building and quizzes, 2) corrective feedback on individual book reports (not only to give feedback but also to give advice on text selection and to encourage reading) during their sustained silent reading, and 3) reading exercises to promote fluent reading without regression, such as quick phrase-by-phrase translation, reading-while-listening (a practice for developing good reading habits by eliminating pointless regressive eye movements),

repeated reading, reading aloud, or shadowing, 4) oral presentation in English - story retelling in pairs. Among the four classes, however, there were some variations of activities depending on students' English proficiency.

PPTs and ERTs were administered as a pretest in April and as a post-test in December. The participants were given 20 minutes for PPTs, considering their concentration and the difficulty of the latter parts of passages, though EPER expects us to allot 60 minutes for PPTs. The average number of blanks filled in was 54 out of 140 in 20 minutes in the pretest. A correct answer was scored 1. On the other hand, 30 minutes were allotted for ERTs. According to the answer key, in spite of there being 20 questions in both version 1 and 2 of Level D and Level F, the maximum possible score is 30. Many correct answers were scored one, but the other correct answers were weighted and given two points.

2.4 *Experimental design*

A pre/post-test design was used for the experimental group in this study. It is ideal to have a few control groups in an experimental study (Nakanishi, 2015), to be sure, but it is impossible to set a control group, because all the participants in the Department of Global English are involved in the Extensive Reading course as a required subject. Thus, it is very difficult to create true experimental conditions in real educational settings (Grabe, 2004).

A counterbalance design was adopted for ERTs. ERT Level D was used for A and B classes, ERT Level F for C and D classes. Each level has two kinds of versions: Version 1 and Version 2. It is inappropriate to use the same reading comprehension test both as a pretest and as a post-test, because the content of a long narrative text can be remembered for some time. As Table 2 shows, Version 1 and Version 2 were counterbalanced: half of A & B class participants took Version 1 for the pretest, and the other half had Version 2, and vice versa for the posttest.

Table 2
Counterbalanced Measures Design for Version 1 and Version 2 of EPER Extensive Reading Tests

	Pretest	Posttest
Group 1 of A & B classes	Version 1	Version 2
Group 2 of A & B classes	Version 2	Version 1
Group 1 of C & D classes	Version 1	Version 2
Group 2 of C & D classes	Version 2	Version 2

Davies and Irvine (1996) demonstrated that parallel forms reliability (a measure of reliability obtained by administering different versions of an assessment tool to the same group of individuals) for each level of the Extensive Reading Test were from 0.92 to 0.99. Since both versions cannot be given to the same participants at the same time in a pre/post-test design, *t*-tests were conducted to prove there was no difference in text difficulty between Version 1 and Version

2. *P*-values (0.61 for D-level reading test, 0.99 for F-level reading test) in Table 3 demonstrate that the two versions are not statistically different. The results of Version 1 and Version 2 were combined as the results of the pretest for statistical analysis.

Table 3
T-test on the differences of text difficulty between Version 1 and Version 2 Based on D Level and F Level Extensive Reading Pre-Tests

	MPS	Version 1		Version 2		*t*-value	*p*-value
		M	*SD*	*M*	*SD*		
D level reading test	30	9.91	5.75	9.04	4.88	.518	.61
F level reading test	30	12.00	4.66	11.98	6.36	.014	.99

MPS=Maximum Possible Scores

3. Results

Table 4 summarizes the descriptive statistics of pre- and post- placement/progress tests. The mean scores of the post-tests in all groups are higher than those of the pretests. In order to examine the differences between the pretests and the post-tests, repeated measures *t*-tests were computed. As table 5 shows, the differences were statistically significant for A class, $t(29) = -5.67$, $p < .01$, $d = 0.68$; for B class, $t(25) = -5.67$, $p < .01$, $d = 0.88$; for C class, $t(25) = -2.38$, $p < .05$, $d = 0.51$; for D class, $t(19) = -2.66$, $p < .05$, $d = 0.49$; for the whole participants, $t(29) = -7.76$, $p < .01$, $d = 0.56$.

Table 4
The Descriptive Statistics of Pre- and Post- Placement/progress Tests

Groups	Pretest			Posttest		
	M	*SD*	Range	*M*	*SD*	Range
A Class (*n*=29)	31.90	11.11	13 - 64	40.41	13.85	9 - 81
B Class (*n*=25)	22.64	6.90	9 - 36	30.28	10.08	18 - 81
C Class (*n*=25)	22.32	8.24	10 - 46	27.44	11.24	10 - 49
D Class (*n*=19)	17.95	6.89	8 - 36	22.37	10.46	9 - 42
Total (*N*=98)	24.39	9.99	8 - 64	31.02	13.29	9 - 81

Table 5
Repeated measures t-tests on Score Development between Pre- and Post- Placement/progress Tests

Pairs	*M*	*SD*	*SE*	*t*-value	Cohen's *d*
A Class pre-test・post-test (*n*=29)	-8.52	8.09	.85	-5.67**	0.68
B Class pre-test・post-test (*n*=25)	-7.64	6.74	1.50	-5.67**	0.88
C Class pre-test・post-test (*n*=25)	-5.12	10.78	1.35	-2.38*	0.51
D Class pre-test・post-test (*n*=19)	-4.42	7.26	2.16	-2.66*	0.49
Total pre-test・post-test (*N*=98)	-6.63	8.46	1.67	-7.76**	0.56

** $p < .01$, * $p < .05$

Since Level D reading tests were used for A and B classes and Level F reading tests for C and D classes, two separate tables were made. Table 6 summarizes the descriptive statistics of pre- and post- extensive reading tests for A and B groups. The mean scores of the post-tests for A and

B groups are much higher than those of the pretests. In order to examine the differences between the pretests and the post-tests, repeated measures t-tests were conducted. As Table 7 shows, the differences were statistically significant for A class, $t(25) = -8.04$, $p < .01$, $d = 1.31$; for B class, $t(25) = -11.81$, $p < .01$, $d = 3.17$; for the whole participants, $t(50) = -13.00$, $p < .01$, $d = 1.81$.

Table 6

Descriptive Statistics of Pre- and Post- Extensive Reading Tests

Groups	Pre-test			Post-test		
	M	SD	Range	M	SD	Range
A Class (n =25)	12.24	5.40	4 - 25	19.30	5.30	11 - 28
B Class (n =25)	6.28	2.70	2 - 13	17.30	4.10	5 - 20
Total (N =50)	9.26	5.19	2 - 25	18.33	4.79	5 - 28

Table 7

Repeated Measures t-tests on Reading Development between Pre- and Post- Reading Test Based on EPER D Level Extensive Reading Test

Pairs	M	SD	SE	t-value	Cohen's d
A Class pre-test · post-test (n =25)	-7.06	4.39	.88	-8.04**	1.31
B Class pre-test · post-test (n =25)	-11.08	4.69	.94	-11.81**	3.17
Total pre-test · post-test (N =50)	-9.07	4.93	.70	-13.00**	1.81

** $p < .01$, * $p < .05$

Table 8 summarizes the descriptive statistics of pre- and post- extensive reading tests for A and B groups. The mean scores of the post-tests in A and B groups are higher than those of the pretests. In order to examine the differences between the pretests and the posttests, repeated measures t-tests were conducted. As Table 9 shows, the differences were statistically significant for C class, $t(27) = -6.04$, $p < .01$, $d = 1.18$; for D class, $t(21) = -3.03$, $p < .05$, $d = 0.61$; for all participants, $t(48) = -6.38$, $p < .01$, $d = 0.89$.

Table 8

Descriptive Statistics of Pre- and Post- Extensive Reading Tests

Groups	Pre-test			Post-test		
	M	SD	Range	M	SD	Range
C Class (n =27)	12.43	5.21	3 - 23.5	18.54	5.13	2 - 28
D Class (n =21)	11.93	6.05	1 - 21.5	15.74	6.47	2 - 24
Total (N =48)	12.21	5.54	1 - 23.5	17.31	5.86	2 - 28

Table 9

Repeated Measures t-tests on Reading Development between Pre- and Post- Reading Test Based on EPER F Level Extensive Reading Test

Pairs	M	SD	SE	t-value	Cohen's d
C Class pre-test · post-test (n =27)	-6.11	5.25	1.01	-6.04**	1.18
D Class pre-test · post-test (n =21)	-3.81	5.77	1.26	-3.03*	0.61
Total pre-test · post-test (N =48)	-5.10	5.55	.80	-6.38**	0.89

** $p < .01$, * $p < .05$

4. Discussion

RQ1 asked if ER instruction for 10 months would improve overall English language proficiency. The analysis of PPTs (placement/progress tests) revealed that overall English language ability or lower-level linguistic ability in all classes improved significantly, contrary to Yamashita's study. This is probably because this program (30 weeks) was longer than Yamashita's study (15 weeks). RQ2 asked if ER practice for 10 months develops general reading ability. The results of ERTs (extensive reading tests) revealed that extensive reading ability or general reading comprehension ability in all classes improved significantly. Although reading rates were not measured, insofar as fluency in reading is defined as the ability to read with accuracy and speed (Carver, 1997; Kame'enni & Simmons, 2001), the gains on the test scores can be interpreted as the development of fluency in reading, because ERTs were administered in a limited time (30 minutes) for both pretests and post-tests. It may be worth pointing out, in passing, that the effect size of B class (Cohen's d value is 3.17) is too large, considering that Cohen classified effect sizes as *small* ($d=0.2$), *medium* ($d=0.5$), and *large* ($d \geq 0.8$) (Sullivan & Feinn, 2012). The D level of ERT might be too difficult for B class students in particular as well as A class students who were not accustomed to reading long texts before the extensive reading instruction. It may safely be said that those students got used to reading, and felt confident about reading long, relatively easy texts through exposure to a large amount of print. As for RQ3, ERTs displayed larger improvement than PPTs, as Cohen's d values showed. As Yamashita (2008, p. 661) suggests, "the effects of extensive reading might be manifested more quickly in general reading skills than in L2 linguistic ability, at least for adult L2 learners."

This study demonstrated the effectiveness of extensive reading practice. This achievement was mainly because of students' constant exposure to print; they read one graded reader or more a week for 30 weeks, and in terms of the number of words, 100,000 - 250,000 words on average were read depending on classes. In this study, however, we couldn't carry out a correlational analysis between the improvement of general reading ability/overall English ability and the number of words the participants read, because the four instructors did not have a unified or consistent way of recording or targeting the amount of reading.

The disadvantage of the present test is that beneficial effects might be explained by other factors. Other English classes, such as Intensive Reading, Listening, or Writing, might have exerted a beneficial influence on the results of post- ERTs and post-PPTs. Actually, however, it is impossible to exclude these factors, since those classes are required courses for the students of the Global English Department. Therefore, to diminish the contribution of other courses, we tried to use a test type suitable for assessing extensive reading ability, that is, a relatively long narrative

text in complete agreement with graded readers they had been reading.

In scoring both PPTs and ERTs, we followed the EPER answer key. A partial credit method is plausible for scoring EPER extensive reading tests (ERTs). Half of the items are given two points and it is likely that one point can be given to partially correct responses of those two-point items. Yoshizawa (2014) concluded that "crediting partially correct answers does not provide ordered thresholds for the items with three categories." In the next study, we need to examine the scoring of the ERTs and to have unified criteria.

5. Conclusion

The present study demonstrated the beneficial effects of our extensive reading program in terms of general reading ability and overall language proficiency, and made a small but valuable contribution to shedding light on the importance of extensive reading in English education for adult learners. As a further study, however, we will need to examine the role of motivation in extensive reading, the relationship between reading development and reading, and the effects of extensive reading on other aspects of English ability, such as vocabulary, writing, listening, spelling, collocation and grammar.

References

Carver, R. (1997). Reading for one second, one minute, or one year from the perspective of reading theory. *Scientific Studies of Reading, 1*, 3-43.

Davies, A. & Irvine, A. (1996). Comparing test difficulty and text readability in the evaluation of an extensive reading programme. In M. Milanovic (Ed.), *Studies in language testing 3: Performance testing, cognition and assessment* (pp. 165-183). Cambridge: Cambridge University Press.

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

EPER (Edinburgh Project on Extensive Reading). (1992). *The EPER guide to organizing programmes of extensive reading*. Institute for Applied Language Studies, University of Edinburgh.

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

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

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

Grabe, E., & Stoller, F.L. (2002). *Teaching and researching reading*. Harlow, UK: Pearson Education.

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

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

Kame'enni , E. & Simmons, D. (2001) Introduction to this special issue: The DNA of reading fluency. *Scientific Studies of Reading, 5*, 203-10.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. New York: Longman.

Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal 73*, 440-464.

Krashen, S. D. (1993). *The power of reading: Insights from the research*. Eaglewood, CO: Libraries Unlimited.

Mikulecky, B. S. (1990). *A short course in teaching reading skills*. Reading, MA: Addison Wesley.

Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly, 49*, 6-37.

Nuttall, C. (2005). *Teaching reading skills in a foreign language*. Oxford: Macmillan Education.

Oller, J. W. Jr., & Conrad, C. A. (1971). The cloze technique and ESL proficiency. *Language Learning, 21*, 183-195.

Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching & applied linguistics*. London, UK: Pearson Education.

Samuels, S. J. (1994). Toward a theory of automatic information processing in reading, revisited. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed.) (pp. 816-837). Newark, DE: International Reading Association.

Sullivan, G. M. & Feinn, R. (2012). Using effect size—or why the *p* value is not enough. *Journal of Graduate Medical Education, 4* : 279–282.

Tanaka, H. & Stapleton, P. (2007). Increasing reading input in Japanese high school EFL classrooms: An empirical study exploring the efficiency of extensive reading. *The Reading Matrix*, 7, 115-131.

Tono, Y. (2015). *Hasshinryoku wo tsukeru atarashii eigo goishido (A new English vocabulary instruction to develop productive ability)*. Tokyo: Sanseido.

Urquhart, S. & Weir, C. (1998). *Reading in a second language: Process, product and practice*. London: Longman.

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

Yoshizawa, K. (2014). Edinburgh project on extensive reading (EPER) reading comprehension tests: Scoring and setting cutoff scores. 『関西大学外国語学部紀要』 *10*, 33-43.