L2 Reading Proficiency Improvement and Reading Behaviors among Different Levels of Readers: Extensive Reading in an EFL Context

Kiyomi Yoshizawa, Kansai University

Atsuko Takase, Kwansei Gakuin University

Kyoko Otsuki, Nara Prefectural University

This study aims to investigate whether extensive reading (ER) facilitates the development of reading comprehension among English as a foreign language (EFL) Japanese learners, and to understand the reading behaviors of learners whose reading proficiency improvement was lower than others after a one-year ER program. A total of 196 Japanese EFL learners in one-year ER programs participated in the study, and 152 Japanese EFL learners in intensive reading programs participated as a control group. A mixed between/within-subject analysis of variance was conducted to assess the impact of ER. The reading logs of ER participants were analyzed to examine reading behaviors at different reading proficiency levels. The results indicated that the ER group improved their reading proficiency significantly after a one-year ER program. However, the control group showed no such improvement. Patterns of successful and unsuccessful readers emerged through a close examination of the ER participants' reading logs.

The last 30 years have seen an increase in the number of studies involving extensive reading (ER) (Jeon & Day, 2016). Such an increase indicates growing interest in ER as an important component of second- and foreign-language curricula worldwide. The majority of previous studies investigating the effects of ER used a quantitative approach. The number of studies taking the qualitative approach was limited. The current study examined the effects of ER taking a mixed methods approach and attempted to provide practical suggestions for ER practitioners.

Yoshizawa, K., Takase, A., & Otsuki, K. (2020). L2 reading proficiency improvement and reading behaviors among different levels of readers: Extensive reading in an EFL context. *Extensive Reading World Congress Proceedings, 5,* 1-10.

Jeon and Day (2016) conducted a meta-analysis to examine the effectiveness of ER on reading proficiency (i.e., reading comprehension, reading rate, and vocabulary). They examined 49 primary studies published from 1980 to 2014, including studies incorporating the five characteristics of ER described in Day and Bamford (2002, pp. 136-141). The selected studies were grouped into two categories based on research design: studies with experimental-versus-control contrasts and studies with pre-to-posttest contrasts. Both types of studies showed small to medium effect sizes. Jeon and Day further examined whether the effects of ER differed significantly due to moderator variables. They analyzed the effects of moderator variables, including setting (i.e., English as a second language (ESL) setting and EFL setting), age, library size, length of

treatment, text type (i.e., web-based books vs. paper books), ER form, and focus skills. They found that age, setting, text types, and ER forms led to different effects.

While the overall supremacy of the ER approach over traditional approaches to reading was reported by Jeon and Day (2016), studies that were not included in their meta-analysis addressed issues in their respective educational contexts. One of these issues is the reading behaviors of participants in ER programs. Nishizawa et al. (2011) and Nishizawa, Yoshioka, and Nagaoka (2017) examined the reading behaviors of reluctant learners of English who had ER sessions in their school curriculum. In particular, Nishizawa et al. (2011) examined critical factors for a successful ER program based on their nine-year ER practice with technology majors in Japan. Participants in their study were learners at a national institute of technology (NIT) specialized in educating engineers in Japan. The participants had ER classes for two and a half years. By analyzing students' reading amounts and Test of English for International Communication (TOEIC) scores, the authors found that the amount of reading is directly proportional to the increase in TOEIC scores. Based on the amount of reading, they divided the students into seven groups and analyzed them. The means of the TOEIC scores increased as their reading amounts increased. In addition, the authors point out that the reading material should be easy enough for students to process the meaning without having to translate it into their L1. This aspect is significantly important for learners to see reading not as a mechanical drill that requires translation of L2 into L1 but rather as an enjoyable activity so that they can continue to read in L2.

Similarly, Nishizawa et al. (2017) examined the reading amounts and readability levels of English text read by Japanese EFL learners who took ER lessons with traditional reading classes for two and a half years. These participants were NIT students. The authors suggested that the students should read more than 600,000 words before their English proficiency could improve. They also examined the reading logs of nine students and reported that students who had read the easiest books, i.e., graded readers with less than 300 headwords, scored higher on an English proficiency test, while those who had read more difficult books, i.e., graded readers with 300 to 1000 headwords, scored lower. The researchers suggested that NIT students should read 200,000 words from easy-to-read books whose *Yomiyasusa* Level (a readability scale for Japanese EFL learners; henceforth, YL) is less than 1.1.

Nishizawa et al. (2011) and Nishizawa et al. (2017) observed a group of learners who read more than the median number of words read by the other ER groups, although the number of books read was much smaller than the median number of books read by the other ER groups. These students had selected longer and harder-to-read books. Their average TOEIC score was much lower than those of other ER groups in their studies.

Mayuzumi and Miyatsu (2012) also focused on the reading behaviors of university students in a one-year ER program in Japan. The authors observed that some learners increased their scores on a proficiency test (i.e., TOEIC test) at the end of a one-year ER program, while others lowered their scores, although the reading amounts were the same. Mayuzumi and Miyatsu named the former "successful group" and the

latter "unsuccessful group." They analyzed the reading logs of 10% of the learners in each group. In their analyses, reading behaviors of successful and unsuccessful readers emerged. The "successful group" started to read books they could understand and continued to read at a steady pace. In addition, this group was engaged in reading and enjoying the content of the book. On the other hand, the "unsuccessful group" started to read books whose readability levels were higher than their reading levels without reading easy-to-read books sufficiently. Consequently, the comments written by the unsuccessful group showed that they did not understand the content of the book.

Nishizawa et al. (2011), Nishizawa et al. (2017), and Mayuzumi and Miyatsu (2012) indicate that the reading behaviors of ER participants can determine whether learners in an ER program are successful or not and that their reading logs are a valuable source that shows their reading behaviors. The findings of these three studies provide educational implications of great importance to ER practitioners. On the other hand, all three studies used TOEIC tests as a measure to assess the effects of ER. However, a question remains as to whether the inference on learners' reading ability based on TOEIC tests is valid when the target language use domain and test tasks are mainly business-oriented. Moreover, the number of reading logs analyzed by Nishizawa, et al. (2017) and Mayuzumi and Miyatsu is small; a larger sample size is preferable. Therefore, this study aims to examine the following two research questions:

> *1. Does ER have a facilitative effect on the development of reading comprehension among Japanese EFL learners?*
>
> *2. What are the reading behaviors of learners whose reading proficiency improvement was lower than others after a one-year ER?*

Methods

Participants

Participants in this study were 196 Japanese EFL learners in a one-year ER program. The control group had 152 Japanese EFL learners in a one-year intensive reading program. Participants were freshmen and sophomore students at a university located in western Japan. These participants had at least six years of formal English education in Japan prior to university admission. Their English proficiency levels ranged from approximately CEFR A1 to B2. The ER group engaged in sustained silent reading for approximately half of each class time and was required to read outside of class. Each time students finished reading a book, they recorded in their reading log the date, title, series, level, word count, reading time, and a brief comment. The control group had textbook-based classes focusing on reading strategies and appreciation of literary works.

Materials

The Edinburgh Project on Extensive Reading Placement/Progress Test (henceforth "EPER_PPT") is a 141-item cloze test developed as a part of the Edinburgh Project on ER (henceforth "EPER"). Based on the EPER_PPT results, the participants' initial reading levels were determined, and ER appropriate for each level was recommended to the participants in the ER program. The eight reading levels are G, F, E, D, C, B, A, and X, where G is elementary level and X is advanced level. This study used the Edinburgh Project on Extensive Reading: Test of Extensive Reading

(henceforth "EPER reading comprehension test"), a reading comprehension test constructed for each of the eight levels mentioned above (Davies & Irvine, 1996) with two versions available. Levels E and D of Version 2 were used for this study. Level E test has 18 items and the Level D test has 20 items. All items require test takers to write short answers except for one that has a multiple-choice format. To answer the second research question, the reading logs that the ER class participants kept throughout the one-year program were analyzed. Based on Mayuzumi and Miyatsu (2012), each participant's reading log was analyzed on the basis of two aspects—how the participant read ER books throughout the year and the nature of the comments the participant wrote after finishing each book.

Procedures

At the beginning of the course, the EPER_PPT was administered as a pre-test to determine students' English proficiency. The same test was administered at the end of the first semester as a mid-test, and at the end of the second semester as a post-test. Similarly, the EPER reading comprehension test was administered at the beginning and end of the course.

Results and Discussion

Data Analysis

To answer our first research question, Rasch analysis was conducted on the EPER reading comprehension test to create an interval scale for placing items (i.e., test questions) and persons (i.e., test takers). The software WINSTEPS (WINSTEPS, Inc.) was used and the dichotomous model was applied. To confirm that the data fit the model, the following points were examined: item-total statistics, fit of items to the model, fit of persons to the model, targeting of the scale, and reliability. Cronbach's alpha of the EPER reading comprehension test was 0.95. After each analysis, the Rasch programs calculated each item's difficulty and each person's ability. Item difficulties and abilities were used in subsequent analyses.

To answer our second research question, the reading logs of the ER group participants were analyzed. Participants in the ER group were divided into three levels (F, E, and D) based on the results of the EPER_PPT at the beginning of the program. 80 participants were assigned to Level F, 73 to level E, and 44 to Level D. Table 1 shows the number of learners in each group. Unsuccessful readers are those whose EPER reading comprehension test gain scores (i.e., differences between test scores) were zero or negative.

Table 1. Number of Successful and Unsuccessful Readers at each of the three levels

ER Levels	Unsuccessful Readers	Successful Readers	NA	Total
Group F (Low)	10	12	58	80
Group E (Middle)	10	9	54	73
Group D (High)	8	9	27	44
NA	0	0	3	3
Total	28	30	142	200

Note: NA refers to the learners not included in the qualitative analysis of the study.

Table 2. Descriptive Statistics of the EPER Reading Comprehension Test (Logits)

	Time 1		Time 2	
	ER	CONTROL	ER	CONTROL
Means	0.15	-0.07	1.14	-0.05
SD	1.11	1.21	1.21	1.27

Successful readers are those among the top 9 to 12 students in each group.

Results of the Quantitative Analyses

In Rasch analysis, item difficulty and person ability are expressed in logits (log odds ratio). When a person's ability matches the difficulty level of an item, there is a 50% chance of the person answering it correctly. As the difference between person's ability and item difficulty grows in favor of the person, the probability of success of the person also increases. On the other hand, as the difference between person's ability and item difficulty grows in the favor of the item, the probability of failure of the person increases.

Table 2 shows the descriptive statistics at the beginning of the program (Time 1), revealing that the ER group's reading comprehension was slightly better than that of the control group. At the end of the program (Time 2), the ER group showed a nearly one-logit increase in reading comprehension. One the other hand, the control group showed only a slight increase.

A mixed between/within-subject analysis of variance was conducted to assess the impact of ER on participants' reading comprehension. The results of multivariate analyses showed a significant main effect of Time, Wilk's lambda = 0.78, F (1, 335) = 93.47, p < 0.001, and partial eta squared = 0.22.

The results also showed a significant interaction effect between Time and Group, Wilk's lambda = 0.80, F (1, 335) = 86.63, p < 0.001, and partial eta squared = 0.21. This indicates that the effect of Time differed across groups. When a significant interaction was found, simple main effect analyses were conducted to examine the effects of one independent variable (categorical

Table 3. Results of the Simple Main Effect Analyses

Simple effects	Factor	Levels	df	F	p
Simple effect of Group within levels of Time					
Group	Time	Pre	1	3.15	n.s.
		Post	1	77.85	p < 0.001
Simple effect of Time within levels of Group					
Time	Group	Control	1	0.81	n.s.
		ER	1	206.36	p < 0.001

predictor variable) at each level of another independent variable (Takehara, 2013). Table 3 shows the results of the simple main effect analyses.

The results show that there was no significant difference between the Control and ER groups at Time 1. However, there was a significant difference between the groups at Time 2. The ER group outperformed the control group at Time 2. The reading performance between Time 1 and Time 2 was not significant for the control group. On the other hand, the ER group showed a significant difference between Time 1 and Time 2.

Results of the Qualitative Analysis

Table 4 summarizes the findings of a careful examination of learners' reading logs. First, the levels of books that learners read at the beginning of the ER program appear to play a significant role in improving their reading proficiency. Some learners whose reading proficiency had not improved began reading books with a higher rating than the levels prescribed by the EPER_PPT results. This tendency is observed in learners whose reading proficiency has not improved in all groups.

Second, it is noteworthy that learners who did not show improvement read books with inconsistent levels and series. For example,

Table 4. Summary of the Findings of the Reading Logs Examination

	Improved	Not improved
Group F (Low)	Levels of books low enough at the beginning Raise level steadily (consistently) Keep reading easier books Range of series limited	Levels of books NOT low enough at the beginning Neither series of books nor levels (YL) consistent (Levels raised and lowered) Comprehension only through illustration
Group E (Middle)	Levels of books low enough at the beginning Raise the levels of books gradually Keep reading rather easy books Series of books consistent (Read several books from the same series)	Levels of books NOT low enough at the beginning Neither series of books nor levels (YL) consistent (Levels raised and lowered, e.g., YL0.5<=>2.0) Read too fast Simple comments
Group D (High)	Read fully comprehensible books at the beginning Kept reading during the summer vacation Comments show participants' deep involvement in reading (The comments focus on the storyline without paying attention to their own reading behaviors.) Enjoyed reading	Levels of books NOT low enough at the beginning Neither series of books nor levels (YL) consistent (Levels raised and lowered, e.g., YL 0.5<=>2.0) Reading speed was fast from the beginning, or very slow after 10 months Simple comments

for five days, one learner read six books whose levels were Oxford Reading Tree (ORT) YL 0.5, Foundations Reading Library (FRL) YL 1.1, Penguin Kids (PGK) YL 0.6, Penguin Readers (PGR) YL 0.8, All Aboard Reading (AAR) YL 1.3, and FRL YL 1.1.

Third, it appears that, for learners whose reading proficiency had not improved, reading comprehension occurs mainly through illustrations found in books. This is revealed in the comment section of the reading logs. Examples of these comments include "This is a story about X" and "David Beckham is quite handsome." These comments seem to show that there was no mental involvement in the stories. These readers simply follow story lines, not stories.

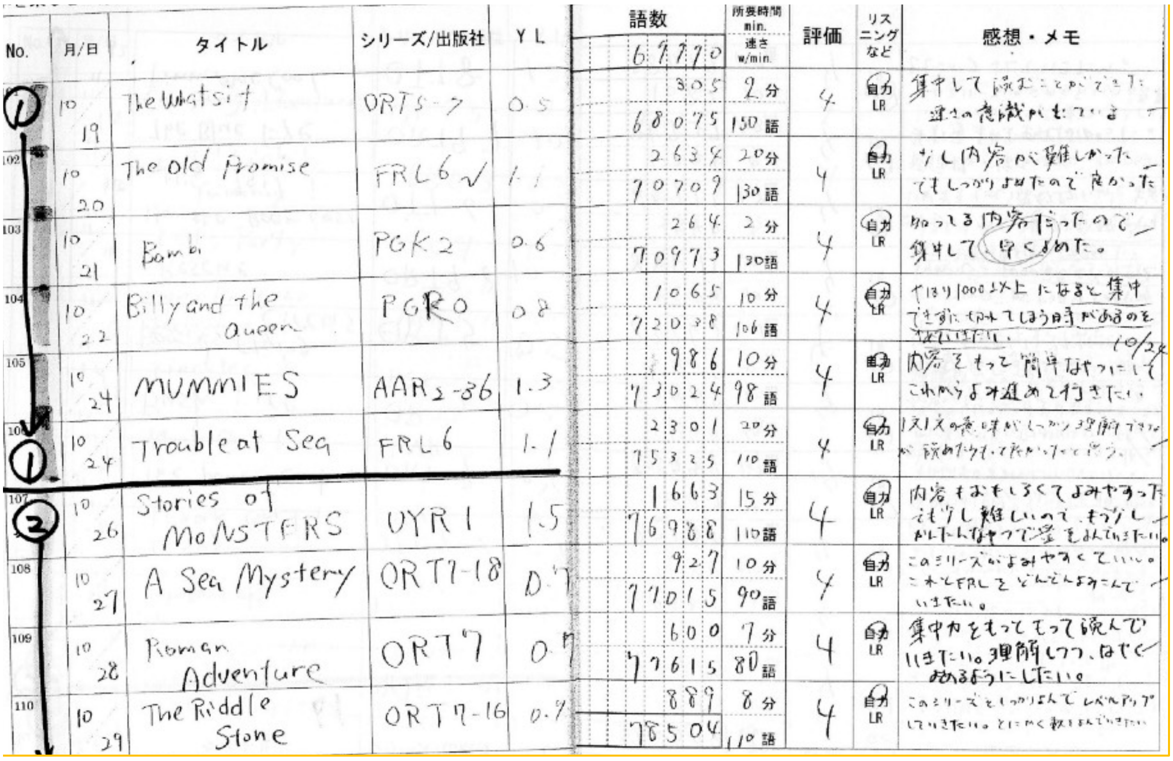
Moreover, many of these learners also wrote comments based on their own reading activities, such as "I could read the book at a good speed," and "the word count of the book is high, but it is not hard to read as there are numerous conversational sentences." They were quite objective about their reading and focused on how fast they read, how they focused on reading, and how they could achieve the amount of reading within the defined dates. It appears that they simply paid attention to their reading behavior without appreciating the stories.

The second and third features mentioned above are illustrated in Figure 1 below, which is a reproduction of a learner's reading log.

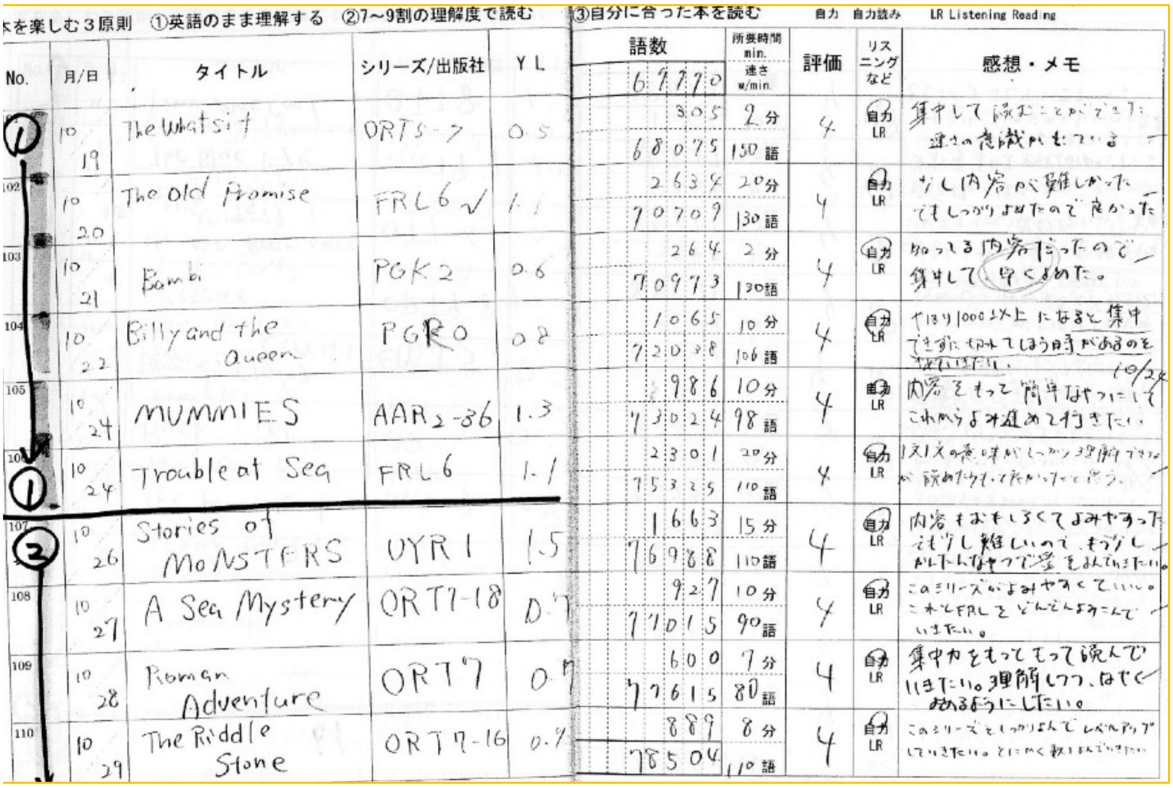
The fifth column from the left (YL) shows the readability level of the books each student has read. The fluctuation in column YL is evident: 0.5 -> 1.1 -> 0.6 -> 0.8 -> 1.3 -> 1.1 -> 1.5 -> 0.7 -> 0.7 -> 0.7. The range of YL is rather wide considering the period during which the learner read these books (11 days, from October 19 to 29). Moreover, the log shows that the book series are not consistent. The fourth column from the left presents book series (in the form of abbreviation) that learners read. Series such as ORT and FRL include recurring characters that can serve to familiarize readers with story events as if these events were happening around readers themselves. The learner's arbitrary choice of books may suggest that the learner did not get involved in the stories and simply followed the storylines. Thus, inconsistency in both the readability levels of books and book series has combined with reading behavior in an unfocused and uncertain manner.

Furthermore, the comment column (last column on the right) may indicate that the learner in question focuses on the reading activity. Comments include "I could focus on reading," "what the book says is hard to understand, but I am glad to read it through," and "I would like more concentration for reading." It may be possible to argue that the learner failed to internalize the characters' experiences. This could become apparent when comparing these comments with those of a different learner. Another learner read a book in which a little boy goes to a swimming pool in a swimsuit. The learner wrote comments about how she had the same experience as the boy in the story. Apparently, learners who wrote their own experiences succeeded in constructing a network from information in text and background knowledge of their own.

After understanding the features that seem to hamper reading proficiency improvement, it may be worthwhile to observe attributes that seem to promote reading comprehension. One learner's reading proficiency significantly improved from the pre- to the post-reading comprehension tests. The learner's correct answers



No.	月/日	タイトル	シリーズ/出版社	YL	語数 (67770)	所要時間 min. / 速さ w/min.	評価	リスニングなど	感想・メモ
101	10/19	The Whatsit	ORTS-7	0.5	305 / 68075	2分 / 150語	4	自力 LR	集中して読むことができた。速さの意識がもてている
102	10/20	The Old Promise	FRL6 ✓	1.1	2634 / 70709	20分 / 130語	4	自力 LR	少し内容が難しかった。でもしっかりよめたので良かった
103	10/21	Bambi	PGK2	0.6	264 / 70973	2分 / 130語	4	自力 LR	知ってる内容だったので集中して早くよめた。
104	10/22	Billy and the Queen	PGR0	0.8	1065 / 72038	10分 / 106語	4	自力 LR	やはり1000以上になると集中できずに切れてしまう時があるのをなおしたい。
105	10/24	MUMMIES	AAR2-86	1.3	986 / 73024	10分 / 98語	4	自力 LR	10/24 内容をもって簡単なやつにしてこれからよみ進めて行きたい。
106	10/24	Trouble at Sea	FRL6	1.1	2301 / 75325	20分 / 110語	4	自力 LR	1文1文の意味がしっかり理解できたが読めた[illegible]
107	10/26	Stories of MONSTERS	UYR1	1.5	1663 / 76988	15分 / 110語	4	自力 LR	内容もおもしろくてよみやすかった。でも少し難しいので、もう少しかんたんなやつで量をよんでいきたい。
108	10/27	A Sea Mystery	ORT7-18	0.7	927 / 77015	10分 / 90語	4	自力 LR	このシリーズがよみやすくていい。これとFRLをどんどんよみこんでいきたい。
109	10/28	Roman Adventure	ORT7	0.7	600 / 77615	7分 / 80語	4	自力 LR	集中力をもってもって読んでいきたい。理解しつつ、はやくよめるようにしたい。
110	10/29	The Riddle Stone	ORT7-16	0.7	889 / 78504	8分 / 110語	4	自力 LR	このシリーズをしっかりよんでレベルアップしていきたい。とにかく数をよんでいきたい。



本を楽しむ3原則 ①英語のまま理解する ②7～9割の理解度で読む ③自分に合った本を読む 自力 自力読み LR Listening Reading

No.	月/日	タイトル	シリーズ/出版社	YL	語数 (67770)	所要時間 min. / 速さ w/min.	評価	リスニングなど	感想・メモ
101	10/19	The Whatsit	ORTS-7	0.5	305 / 68075	2分 / 150語	4	自力 LR	集中して読むことができた。速さの意識がもてている
102	10/20	The Old Promise	FRL6 ✓	1.1	2634 / 70709	20分 / 130語	4	自力 LR	少し内容が難しかった。でもしっかりよめたので良かった
103	10/21	Bambi	PGK2	0.6	264 / 70973	2分 / 130語	4	自力 LR	知ってる内容だったので集中して早くよめた。
104	10/22	Billy and the Queen	PGR0	0.8	1065 / 72038	10分 / 106語	4	自力 LR	やはり1000以上になると集中できずに切れてしまう時があるのをなおしたい。
105	10/24	MUMMIES	AAR2-86	1.3	986 / 73024	10分 / 98語	4	自力 LR	10/24 内容をもって簡単なやつにしてこれからよみ進めて行きたい。
106	10/24	Trouble at Sea	FRL6	1.1	2301 / 75325	20分 / 110語	4	自力 LR	1文1文の意味がしっかり理解できたが読めた[illegible]
107	10/26	Stories of MONSTERS	UYR1	1.5	1663 / 76988	15分 / 110語	4	自力 LR	内容もおもしろくてよみやすかった。でも少し難しいので、もう少しかんたんなやつで量をよんでいきたい。
108	10/27	A Sea Mystery	ORT7-18	0.7	927 / 77015	10分 / 90語	4	自力 LR	このシリーズがよみやすくていい。これとFRLをどんどんよみこんでいきたい。
109	10/28	Roman Adventure	ORT7	0.7	600 / 77615	7分 / 80語	4	自力 LR	集中力をもってもって読んでいきたい。理解しつつ、はやくよめるようにしたい。
110	10/29	The Riddle Stone	ORT7-16	0.7	889 / 78504	8分 / 110語	4	自力 LR	このシリーズをしっかりよんでレベルアップしていきたい。とにかく数をよんでいきたい。

Figure 1. Example of a reading log produced by a learner whose reading proficiency has not improved.

improved from 18 (47%) in the pretest to 38 (100%) in the posttest. Her EPER_PPT also improved: 66 (47%) in the pretest => 107(76%) in the posttest. The reading log of this student shows the following attributes:

- Book levels low enough at the beginning
- Raised level steadily (consistently)
- Kept reading regularly all year around
- Read various series of books
- Read the same series at a time

The above investigation of reading logs shows that, although ample exposure to text is the nucleus of ER, reading behavior seems to play a vital role in improving learners' reading comprehension.

The current study examined whether ER has a facilitative effect on the reading comprehension among Japanese EFL learners. The results of the quantitative analyses indicate that the ER group improved their reading proficiency significantly after the one-year ER program. However, a similar improvement was not observed in the control group. On the other hand, about 14% of the learners in the ER group did not improve their reading proficiency. Examination of the learners' reading logs shows distinctive reading behaviors that are likely to promote or inhibit reading proficiency improvement.

One of the observed differences between the two groups of learners was that unsuccessful learners did not select books at a sufficiently low reading level at the beginning of the ER program. In some cases, these learners did not read enough easy-to-read books before moving on to books on a par with their reading proficiency. Before the learners in the current study participated in the ER program, it is likely that reading was for learning an L2, and readability level of text was usually higher than their reading proficiency levels. L2-to-L1 translation with the use of a bilingual dictionary was a usual means to comprehend text. In the ER program, reading easy-to-read books before they started to read books appropriate to their reading proficiency was intended to make them break from their old reading habits and gradually read for general comprehension without depending on constant L2-to-L1 translation (Mayuzumi & Miyatsu, 2012; Nishizawa, et al., 2011; Nishizawa, et al., 2017). Another reading behavior distinguishing the successful learners from the unsuccessful ones was that the former kept reading constantly and raised the readability levels of ER books incrementally.

Grabe (2009) describes multiple levels of processes that are involved in fluent and proficient reading: from word to text levels. Information processed at any level becomes active and it forms a network of information that becomes working memory. The capacity of working memory is limited in terms of storage, "linkages to long-term memory," and "abilities to carry out multiple processes simultaneously" (p. 32). Those aforementioned reading behaviors of successful readers (i.e., sufficient amount of reading easy-to-read books and incremental increase of the levels of ER books) contribute to reduce the loads on the learners' working memory while they engage in reading extended text.

Close observation of learners' reading behaviors will help ER practitioners detect unsuccessful reading patterns and guide potential unsuccessful readers in the appropriate direction.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 18K00766.

References

Davies, A., & Irvine, A. (1996). Comparing test difficulty and text readability in the evaluation of an extensive reading programme. In M. Milanovic & N. Saville (Eds.), *Studies in Language Testing 3* (pp. 165-183). Cambridge, United Kingdom: Cambridge University Press.

Day, R., & Bamford, J. (2002). Top Ten Principles for Teaching Extensive Reading. *Reading in a Foreign Language, 14*(2), 136-141.

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

Jeon, E., & Day, R. (2016). The effectiveness of ER on reading proficiency: A meta-analysis. *Reading in a Foreign Language, 28* (2), 246-265.

Mayuzumi, M., & Miyatsu, T. (2012). Tadoku no koka wo takameru ni ha—dokusho keiko no kosatsu kara [What enhances the effect of extensive reading: A consideration about reading attitudes]. *JERA Bulletin, 6*, 60-71.

Nishizawa, H., Yoshioka, T., & Nagaoka, M. (2017). How many words should elementary EFL learners read extensively and from which readability levels. *Toyota Kogyo Senmon Gakko Kiyo* [Journal of Toyota National Institute of Technology], 50, 1-12.

Nishizawa, H., Yoshioka, T., Ito, K., Nagaoka, M., Hiroyama, S., & Asai, H. (2011). Critical factors in a successful extensive reading program of English. *Journal of Japanese Society for Engineering Education, 59*(4), 66-71.

Takehara, T. (2013). *SPSS-no susume 1* [Encouraging the Use of SPSS]. Kyoto: Kitaoji-shobo.