The effectiveness of ER on reading proficiency: A meta-analysis

Eun-Young Jeon
Chonbuk National University
Korea

Richard R. Day
University of Hawai'i
United States

Abstract

A meta-analysis was performed to investigate the impact of extensive reading (ER) on reading proficiency. This study gathered 71 unique samples from 49 primary studies published from 1980 to 2014 involving a total of 5,919 participants. Effect sizes were generated separately according to two different study designs: experimental-versus-control contrasts and pre-to-post-test contrasts. Small to medium effect was found in both study designs. Moderator analysis showed growing interest in ER in the field over the last 30 years. Also, a higher effect was found in the adults than in the children and adolescents group. English as a foreign language (EFL) settings showed a higher effect than English as a second language (ESL) settings; and web-based stories had a higher effect than paper books. Finally, ER as a part of curriculum showed the highest mean effect among ER types. Suggestions are made on how to implement ER in ESL and EFL settings effectively.

Keywords: extensive reading, meta-analysis, reading comprehension, reading rate, vocabulary

Providing rich input in English is essential for promoting English proficiency. Extensive reading (ER) is an excellent way to provide target language input, especially in foreign language settings where the target language input is very limited. ER can be defined as "an approach to language teaching in which learners read a lot of easy material in the new language" (Bamford & Day, 2004, p. 1). In other words, ER is a way of learning a language through a great amount of reading for pleasure. It is sometimes called as "pleasure reading," "free voluntary reading" and "sustained silent reading."

The contribution of ER to various aspects of language proficiency has been confirmed by a number of research studies: vocabulary (Horst, 2005; Zimmerman, 1997); reading comprehension (Elley & Mangubhai, 1983; Robb & Kano, 2013); reading rate (Beglar, Hunt & Kite, 2012; Bell, 2001); and writing (Im, Ahn &Yoon, 2010; Tsang, 1996). Moreover, its effect

on the affective domain such as motivation has also been demonstrated by a number of studies (e.g., Al-Homoud & Schmitt, 2009; Kim & Hwang, 2006). Despite its attested effect on English proficiency, ER is not yet widely practiced in either English as a second language (ESL) or English as a foreign language (EFL) settings. In reality, ER is still "an approach less traveled" (Day & Bamford, 1998, p. 164). According to a survey by Kim (2002), 94% of high-school English teachers in Korea use the grammar-translation method rather than an ER approach for teaching English reading in the classroom.

The aim of this study is to investigate how much effect we can expect from ER intervention through the tool of meta-analysis. In recent years, meta-analysis has gained attention in the field of second language studies with an increasing amount of research in the field and the growing need to integrate the previous findings. Also, computing effect sizes were greatly facilitated by the development of software programs such as Comprehensive Meta-Analysis and RevMan.

In order to investigate overall effectiveness of ER and variables which contribute to this effect, the present study posed the following research questions:

1. What is the overall effectiveness of ER on reading proficiency (reading comprehension, reading rate, and vocabulary) in ESL and EFL settings?
2. To what extent do identification (year of publication), context (age, ESL and EFL setting, library size), treatment (length, text type, ER form), and outcome (reading comprehension, reading rate, and vocabulary) variables affect the impact of ER?[1]

Literature Review

There have been some efforts to integrate findings on ER. Krashen (2007) did a meta-analysis on the effect of ER on EFL adolescents and young adults. ER was found to have a large effect on reading comprehension ($d = 0.88$) and a medium effect on a cloze test ($d = 0.73$). Moreover, a positive relationship was found between the number of books read per student and reading comprehension but no relationship was found between the number of books per student and the results of the cloze test.

Kim (2012) conducted a comprehensive meta-analysis on ER encompassing the cognitive domain as well as the affective domain. The cognitive domain included listening, vocabulary, literacy, reading speed, reading comprehension, and writing; while the affective domain included interest, confidence, motivation, attendance, attitude, and anxiety. Using Hedges' (1981) formula, the mean effect size for the cognitive domain was 0.62 and the affective domain was 0.40.

Nakanishi (2015) explored the overall strength of ER and how its effect differed depending on the participants' ages and periods of instruction. The overall effectiveness of ER was medium for

[1] We used the term "reading proficiency" to describe proficiency in reading comprehension, reading rate, and vocabulary. The rationale for this is based on the wide acceptance of the positive relationship among reading comprehension, reading rate and vocabulary in the field of language acquisition. In fact, a high level of reading comprehension cannot take place without efficient decoding of print (Adams, 1994) and vocabulary knowledge.

both study designs, group contrasts ($d = 0.46$) and pre-post contrasts ($d = 0.71$). Moreover, a larger mean effect was found in the older participants ($d = 0.67$) and in the longer programs ($d = 0.52$).

There were a number of shortcomings in the previous research that motivated the present research. First, the samples were not large enough to claim the validity of the meta-analysis. Krashen (2007)'s meta-analysis included only nine studies for reading comprehension and 14 studies for the cloze test. Kim (2012) had a larger number of studies (i.e., 21 studies) compared to Krashen (2007). This was due to the fact that practically any study on ER was included in his investigation under the terms "cognitive domain" and "affective domain." However, the effect of ER can vary depending on the focus of second language (L2) ability (Yamashita, 2008). For example, the effect of ER on reading comprehension can be different from its effect on writing. Therefore, to obtain a better understanding on effect of ER, it is better to do separate meta-analyses on different aspects of L2 ability rather than combining them all in a single category such as cognitive domain or affective domain. Nakanishi (2015) included the greatest number of samples up to now: 22 samples for group contrasts and 21 samples for pre-post contrasts.

Although Nakanishi's study (2015) included a large number of samples to claim the validity of the meta-analysis, its shortcoming was that a number of studies included in the meta-analysis did not properly reflect characteristics of ER suggested by Day and Bamford (2002), particularly, the self-selection of books. That is, some of the primary studies included in his meta-analysis were more like obligatory assigned reading of long texts (e.g., Kweon & Kim, 2008; Lao & Krashen, 2000; Lin, 2010) rather than free selective pleasure reading. Also, some studies were not specifically targeted for English as a foreign or second language (e.g., Greenberg, Rodrigo, Berry, Brinck, & Joseph, 2006: English as a native language; Rodrigo, Krashen, & Gribbons, 2004: Spanish as a foreign language). Therefore, it is necessary to redo the analysis by narrowing the study samples according to the common conceptualization of ER.

Furthermore, while previous meta-analyses have demonstrated the overall effectiveness of ER in comparison with traditional intensive approaches to reading, there are some unexplored issues related to the implementation of ER. For example, no study has explored how variables such as setting, library size, text type, and ER implementation affect the effectiveness of ER. Compared to previous meta-analyses, this study applied stricter inclusion criteria by only including primary studies which truly reflect ER features suggested by Day and Bamford (2002) in either ESL or EFL settings. In addition, by investigating some unexplored moderator variables, we tried to fill the gaps in the literature by providing answers to the practical issues of ER implementation.

Method

Search of literature

In order to find and select the necessary data for the meta-analysis, exhaustive online and manual bibliographical searches were conducted. The educational databases including Educational Resources Information Center (ERIC), Research Information Sharing Service (RISS), and Dissertation and Thesis, were utilized as tools for the online search. Search terms included

combinations of the following key words: ER, pleasure reading, reading comprehension, reading rate, vocabulary, ESL, and EFL. In addition, seven applied linguistics journals were searched manually: *Reading Research Quarterly, Reading in a Foreign Language, TESOL Quarterly, Language Learning, Applied Linguistics, ELT Journal,* and *System.* Finally, the ER Foundation bibliography was examined which had a list of about 500 references on ER.

Criteria for inclusion

The inclusion criteria were constructed in terms of ER and meta-analysis. In order to be included in the meta-analysis, the individual studies had to embody the following five characteristics of ER from Day and Bamford (2002, pp. 137–140):

    i.The reading material is easy (suitable for learners' levels).
    ii.Learners choose what they want to read.
    iii.Learners read as much as possible.
    iv.Reading is individual and silent.
    v.Teachers orient and guide their students.

While investigating related studies, we noticed that there was a lack of a common understanding of what ER is. For example, many studies which claimed to have adopted an ER approach turned out to be, in fact, assigned readings of long texts. Such studies were excluded in our meta-analysis since they did not properly reflect the self-selected reading principle of ER. This principle is important since students were found to have higher satisfaction with books of their choice over obligatory assigned reading (e.g., Hayashi, 1999; Park & Kang, 2004; Sheu, 2003).

In terms of meta-analysis, studies had to satisfy the following criteria:

1. Studies should adopt either experimental or quasi-experimental design which has quantifiable data for meta-analysis. That is, in order to calculate its effect, a study should have statistical information such as means, standard deviations, number of participants, t-value, p-value and so on. Some primary studies that did not report enough statistical information to compute effect size had to be excluded. Often studies failed to report the standard deviation and those studies had to be excluded (e.g., Krashen, 1989). Results should indicate changes in learners' reading comprehension, reading rate, or vocabulary.

2. The studies from 1980 to 2014 were included because studies on ER in ESL/EFL settings were extremely hard to find prior to 1980. Even if there were, in most cases they did not contain necessary information to compute effect sizes, nor did they adopt principals of ER advanced by Day and Bamford (2002). Thus, only the studies from 1980 onwards were included.

Using the above mentioned criteria, 51 samples from 32 primary studies were selected for experimental- vs. control-group design and 20 samples from 17 primary studies were selected for pre- vs. post- test design.

Coding of study reports

The details of each study were coded as shown in Table 1. First, studies were identified in terms of author, title, year, and type of publication. Then they were coded with respect to context, treatment, and outcome variables according to the research questions. Contextual information included the setting of the studies, participant's age, and library size. Participants' ages were classified into three age groups: children, adolescents, and adults. The children's group involved participants in elementary school, the adolescent's group from middle school to high school, and the adult's group from college level and above. Library size was classified into three categories depending on how many books were made available to students. Some studies which used web stories or did not report the number of available books were excluded. Libraries consisting of fewer than 100 books were coded as small, from 100 to 999 books as medium, and over 1,000 books as large.

Table 1. *Coded data from primary studies*

Variables	Values			
Identification				
Author				
Title				
Year of publication				
Type of publication	Article	Report	Dissertation	
Context				
Second or foreign language setting	SL	FL		
Country				
Participant's age	Children	Adolescents	Adults	
Library size	Small	Medium	Large	
Treatment				
Length	Short	Medium	Long	
Text type	Paper	Web		
Implementation	Type A	Type B	Type C	Type D
Outcome				
Dependent variable	RC	RR	V	
Statistical information of control and experimental Group	N-size	Mean	Standard deviation	Effect sizes

Notes. Type A: ER as an independent reading course, Type B: ER as a part of reading course, Type C: ER as a part of curriculum, Type D: ER as an extracurricular activity. RC= reading comprehension, RR= reading rate, V= vocabulary

Treatment information included length of the program and the type of text used for reading. The length of the reading programs was coded into three categories of long, medium, and short term. Since most studies were carried out in semester units, the investigations carried out for about a semester or less (i.e., < 5 months) were classified as short term; the investigations which lasted up to an academic year (i.e., 6 ~ 10 months) were classified as medium term; and the investigation over one academic year were classified as long-term. The type of text was classified into two categories of paper or web depending on the format of the reading text.

How ER was integrated in the language curriculum was classified in accordance with Day and Bamford (1998, p. 41):
- as a separate, stand-alone course
- as part of an existing reading course
- as a non-credit addition to an existing course
- as an extracurricular activity

Finally, outcome information included test scores for reading comprehension and vocabulary, while words read per minute was coded for reading rate. Specifically, all the necessary statistical information to compute effect sizes (e.g., number of participants, means, standard deviations) was coded.

Meta-analysis

We used the Comprehensive Meta-Analysis Program to calculate the effect sizes. Effect sizes can be calculated in several ways, such as Cohen's *d*, Hedges' *g*, Pearson's *r*, and Glass's *delta*. Cohen's *d* was adopted in this study because it is considered to be the most typical way to estimate effect sizes (Cohen, 1977).

Cohen (1977, 1988) proposed a guideline for interpreting the effect size derived from standardized mean differences based on his observations of the typical range of findings in social science research: small effects ($d \leq 0.2$), medium effects ($0.2 < d < 0.8$), and large effects ($0.8 \leq d$). However, effect sizes are best understood when they are interpreted within a specific domain. Based on the observed effect from 346 primary studies and 91 meta-analyses, Plonsky and Oswald (2014) suggested a new field-specific scale for L2 research. As shown in Table 2, their benchmark provides a higher range than that of Cohen's.

Table 2. *Benchmark for interpreting effect sizes in L2 research* (Plonsky & Oswald, 2014)

		Small	Medium	Large
d	Between group contrast	0.4	0.7	1.0
	Pre- to post-test contrast	0.6	1.0	1.4

The effect sizes based on experimental- and control-group comparisons were meta-analyzed separately from those based on pre- and post-test comparisons (Lipsey &Wilson, 2001).

Results

Publication bias

A funnel plot and Fail-safe N test were employed to check for the publication bias. If the distribution of the studies forms a symmetrical funnel shape, there is no publication bias. However, deviation from this shape may indicate the existence of publication bias and of more variability than expected from simple sampling error. The distribution of the observed studies is more or less symmetrical as shown in Figures 1 and 2.
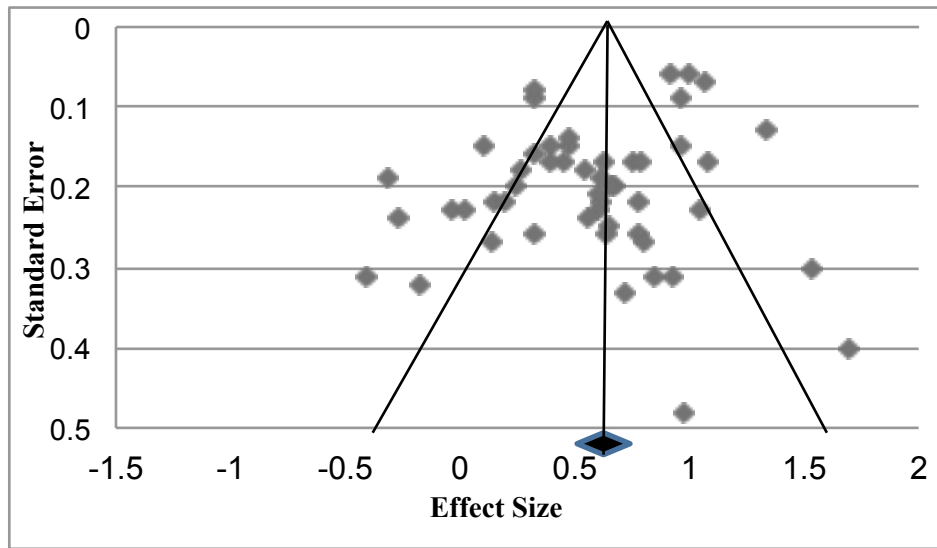
Figure 1. Funnel plot of experimental- vs. control-group comparison.
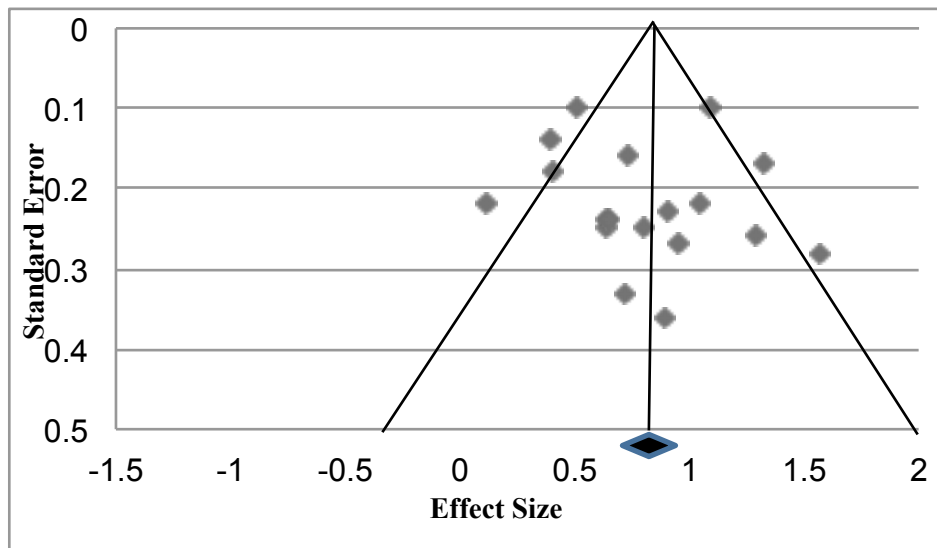*Note*. The diamond indicates the summary effect.

Figure 2. Funnel plot of experimental- vs. control-group comparison.
Note. The diamond indicates the summary effect.

Also the results of Fail-safe N test revealed that unpublished studies reporting non-significant findings are unlikely to reverse the findings. For group comparisons, 8530 missing studies are needed; and for pre-to-post-test comparisons, 1506 studies are needed to bring p-value to $p > .05$ level. Therefore, we can have confidence in the data because it is very unlikely that 8530 and 1506 studies with nil-null results were not included in the meta-analysis because they failed to be published or because they were simply overlooked by the researcher in the literature review.

Overall effectiveness of ER

The first research question concerns the overall effectiveness of an ER approach on reading proficiency. The overall effectiveness of ER was synthesized separately based on the study design (i.e., experimental- vs. control-group design or pre-to-post-test design) as shown in Tables 3 and 4. If groups in a study consisted of different members, they were considered as unique samples and the effect sizes were yielded for each of them. In such cases, more than one sample was drawn from a single study. For example, Lai (1993a)'s study provided seven samples from seven independent schools (i.e., School A ~ School H).

The overall effectiveness aggregated from 51 experimental- vs. control-group contrasts was 0.57. This indicates the superiority of the ER group over the intensive or traditional reading group on the immediate post-test. The confidence interval shows that with 95% certainty the true effect is far from zero, falling anywhere within the medium effect range (between 0.46 and 0.68). Moreover, when the effect size was viewed in terms of the average percentile standing, the mean of the experimental group was at the 72th percentile of the control group. This clearly shows the impact of ER when compared to a traditional reading approach. Finally, the homogeneity test was found to be statistically significant ($Q = 316.80$, $df = 50$, $p < .01$), which indicated that there is more variability in effect sizes than would be expected from sampling error around the mean. In other words, there is evidence for additional systematic sources of variability (i.e., moderators).

Table 3. *Treatment effects for experimental- vs. control- group comparisons*

Study names	N (Exp)	D	SE	95% CI lower	95% CI upper	Outcome
Alavi & Kayvanshekoh (2012)	18	0.63	0.17	0.30	0.96	V
Al-Homoud & Schimitt (2009)	47	0.15	0.22	-0.29	0.59	RC, RR
Beglar et al. (2012)	35	0.93	0.31	0.32	1.53	RC, RR
Bell (2001)	14	1.70	0.40	0.91	2.48	RC, RR
Burrows (2012)	74	0.75	0.17	0.42	1.08	RC
Cha (2009)	10	0.98	0.48	0.04	1.92	RR, V
Chen, Chen, Chen & Wei (2013)	46	0.61	0.22	0.18	1.04	RC
Cho, Kim, & Krashen (2004)	70	0.45	0.17	0.11	0.78	RC, V
Cho & Choi (2008)	28	0.14	0.27	-0.39	0.67	RC
de Morgado (2009)	30	0.32	0.26	-0.19	0.83	RC
Elley (1991) study3	256	0.33	0.09	0.15	0.50	RC, V
Elley (1991) study3	350	0.32	0.08	0.16	0.47	RC, V
Elley & Mangubhai (1983) class 5	70	1.08	0.17	0.75	1.41	RC
Elley & Mangubhai (1983) class 6	64	0.79	0.17	0.46	1.12	RC, V
Jeon (2008)	17	0.72	0.33	0.07	1.37	RC
Kim & Lim (2012)	35	0.65	0.25	0.16	1.14	RC
Lai (1993a) School A	86	0.47	0.15	0.18	0.76	RC, V

Lai (1993a) School B	59	-0.32	0.19	-0.69	0.05	RC
Lai (1993a) School C	83	0.32	0.16	0.01	0.63	RC, V
Lai (1993a) School D	77	0.39	0.17	0.07	0.71	RC, V
Lai (1993a) School E	40	-0.03	0.23	-0.48	0.42	RC, V
Lai (1993a) School G	39	-0.27	0.24	-0.74	0.20	RC
Lai (1993a) School H	36	0.02	0.23	-0.43	0.47	RC
Lee (2007) study 2	67	0.10	0.15	-0.19	0.39	RC, V
Lee (2007) study 3	41	0.54	0.18	0.19	0.89	RC, V
Lituanas, Jacobs, & Renandya (2001)	30	1.53	0.30	0.95	2.11	RC, RR
Mason & Krashen (1997) study 1	20	-0.17	0.32	-0.80	0.46	RC
Mason & Krashen (1997) study 2 University	40	0.60	0.23	0.15	1.05	RC
Mason & Krashen (1997) study 2 Junior college	31	0.85	0.31	0.24	1.46	RC
Mason & Krashen (1997) study 3	40	0.78	0.22	0.35	1.21	RC, RR
Matsui & Noro (2010)	60	0.27	0.18	-0.09	0.62	RC, RR
Mermelstein (2014)	41	1.04	0.23	0.59	1.49	V
Nakashini & Ueda (2011)	20	-0.41	0.31	-1.02	0.20	RC
Rezaee & Nourzadeh (2011)	26	0.67	0.20	0.28	1.06	RC
Robb & Kano (2013) Economics major	543	1.00	0.06	0.88	1.12	RC
Robb & Kano (2013) Business major	587	0.92	0.06	0.80	1.04	RC
Robb & Kano (2013) Law major	546	1.07	0.07	0.93	1.21	RC
Robb & Kano (2013) Foreign language major	254	0.96	0.09	0.78	1.14	RC
Robb & Kano (2013) Science major	100	0.96	0.15	0.67	1.25	RC
Robb & Kano (2013) Technology major	48	0.66	0.20	0.27	1.05	RC
Robb & Kano (2013) Computer major	142	1.34	0.13	1.09	1.59	RC
Robb & Susser (1989)	63	0.62	0.19	0.26	0.98	RC, RR
Sheu (2003)	34	0.78	0.26	0.28	1.28	RC, RR
Shin (2013)	35	0.56	0.24	0.09	1.03	V
Sims (1996) Group A	30	0.80	0.27	0.27	1.33	RC
Sims (1996) Group B	30	0.64	0.26	0.13	1.15	RC
Smith (2006)	51	0.24	0.20	-0.15	0.63	RC
Tanaka & Stapleton (2007)	96	0.40	0.15	0.11	0.69	RC, RR
Tsang (1996)	48	0.60	0.21	0.19	1.01	V

Weitz (2003)	43	0.20	0.22	-0.24	0.63	RC, V
Yamamoto (2011)	33	0.48	0.14	0.21	0.75	V
Overall	4,683	0.57	0.06	0.46	0.68	RC, RR, V

Notes. RC= reading comprehension, RR= reading rate, V= vocabulary

Similarly, the overall effect from 20 pre-test to post-test comparisons was small to medium (d = 0.79). Moreover, the numbers of participants involved in pre-test to post-test comparison studies tended to be smaller than those of experimental vs. control design, which led to twice as high standard error (SE = 0.09) than the experimental vs. control design (SE = 0.06). Furthermore, the homogeneity test was also found to be statistically significant (Q = 68.88, df = 19, p < .01).

Table 4. *Treatment effects for pretest to posttest comparisons*

Study names	N (Exp)	d	SE	95%CI lower	95% CI upper	Outcome
Fujita & Noro (2009)	68	0.32	0.19	-0.05	0.69	RC, RR
Hafiz & Tudor (1989)	16	1.04	0.22	0.61	1.47	RC
Hayashi (1999) beginning level	35	1.29	0.26	0.77	1.80	RC, V
Hayashi (1999) intermediate level	40	0.90	0.23	0.44	1.36	RC, V
Horst (2005)	17	1.57	0.28	1.02	2.11	V
Huang & Liou (2007)	38	0.64	0.24	0.18	1.11	V
Iwahori (2008)	33	0.64	0.25	0.15	1.14	RC, RR
Kim (2014)	249	0.51	0.10	0.32	0.70	RC, RR
Kim & Hwang (2006)	20	0.72	0.33	0.08	1.36	V
Lai (1993b) S1	126	0.39	0.14	0.12	0.67	RC, RR
Lai (1993b) S2	88	1.32	0.17	0.98	1.65	RC, RR
Lai (1993b) S3	52	0.12	0.22	-0.31	0.56	RC, RR
Mason (2003)	30	0.95	0.27	0.41	1.48	RC
Park & Kang (2004)	35	0.80	0.25	0.31	1.29	RC, V
Taguchi et al. (2004)	10	1.44	0.51	0.45	2.44	RC, RR
Takase (2007)	216	1.09	0.10	0.88	1.29	RC
Takase (2009) group 3	36	0.65	0.24	0.18	1.13	RC
Yamashita (2008)	31	0.41	0.18	0.06	0.77	RC
Yang (2010)	79	0.73	0.16	0.41	1.05	RR
Zimmerman (1997)	17	0.89	0.36	0.19	1.60	V
Overall	1,236	0.79	0.09	0.62	0.96	RC, RR, V

Notes. RC= reading comprehension, RR= reading rate, V= vocabulary.

It can be concluded that effectiveness of an ER approach was small to medium with respect to Plonsky and Oswald (2014)'s proposed benchmark for interpreting effect sizes in SLA. That is, experimental groups outperformed control groups on their immediate post-test and there was a small to medium improvement in reading proficiency from the pre-test to post-test. The overall supremacy of ER over intensive or traditional reading approach is consistent with previous meta-

analyses.

The moderator analyses

In order to assess differences among moderators, a mixed-effect model was employed. Moderator analyses were performed in terms of identification, context, treatment, and outcome variables as shown in Table 5. Q-Statistics were employed to find out which variable type was a significant moderator. Moderator analysis was only performed for control vs. experimental group analysis since pre-to post-test contrast had small numbers of effect sizes for aggregation which, as a result, can jeopardize the validity of the analysis.

Table 5. *Moderator analysis for experimental- vs. control group comparison*s

Group	Moderator Variable		*d*	*k*	*SE*	95%CI lower	95%CI upper	*Q*
Identification	Year	1980s	0.83	3	0.17	0.51	1.16	22.87*
		1990s	0.33	16	0.08	0.18	0.48	
		2000s	0.49	14	0.09	0.31	0.66	
		2010s	0.78	18	0.07	0.65	0.92	
Context	Setting	EFL	0.65	37	0.06	0.53	0.77	5.89*
		ESL	0.38	14	0.09	0.20	0.56	
	Age	Children	0.52	6	0.13	0.26	0.78	9.14*
		Adolescents	0.35	15	0.09	0.17	0.53	
		Adults	0.70	30	0.06	0.56	0.81	
	Library size	Small	0.64	4	0.19	0.27	1.01	2.02
		Medium	0.50	17	0.09	0.33	0.67	
		Large	0.72	8	0.13	0.46	0.97	
Treatment	Length	Short	0.51	16	0.01	0.31	0.71	0.49
		Medium	0.59	31	0.01	0.46	0.73	
		Long	0.60	4	0.03	0.25	0.96	
	Text type	Paper	0.47	41	0.05	0.37	0.57	18.48*
		Web	0.89	10	0.09	0.73	1.06	
	ER Form	Type A	0.24	1	0.20	-0.15	0.63	30.90*
		Type B	0.47	38	0.06	0.35	0.58	
		Type C	0.91	11	0.06	0.79	1.03	
		Type D	0.67	1	0.20	0.28	1.06	
Outcome	Focus skills	RC	0.54	46	0.06	0.42	0.65	4.65
		RR	0.83	10	0.14	0.55	1.11	
		V	0.47	17	0.10	0.28	0.66	

Notes. Type A: ER as an independent reading course, Type B: ER as a part of reading course, Type C: ER as a part of curriculum, Type D: ER as an extracurricular activity. *p < 0.05

(a) Identification. Fifty-one (51) samples were grouped into four different periods of publication: 1980s, 1990s, 2000s, and 2010s. Although 1980s had the highest mean effect ($d = 0.83$), increase in effect sizes was shown as the years went by (i.e., 1990s: $d = 0.33$; 2000s: $d = 0.49$; 2010s: $d = 0.78$), and the difference among moderators were found to be statistically significant ($Q = 22.87$, $p < 0.05$). Increase not only in the mean effect of ER but also in the number of studies reflected the development and interest in ER in the field of applied linguistics over the last 30 years

(b) Setting**.** The difference between ESL ($d = 0.38$) and EFL ($d = 0.65$) settings was found to be statistically significant ($Q = 5.89$, $p < 0.05$). Although an ESL setting is generally considered to be a better environment to practice ER than an EFL setting, the present result indicated the promising outlook of ER in EFL settings. However, this result should be interpreted with caution since EFL settings ($k = 37$) involved more studies than ESL settings ($k = 14$), which may have influenced the standard error of the effect size. Moreover, ESL studies did not involve any adult groups which showed higher mean effect than other age groups. It is also worth noting that the majority of EFL and ESL countries included in the meta-analysis are Asian Pacific countries rather than North American or European countries as shown in Figures 3 and 4.
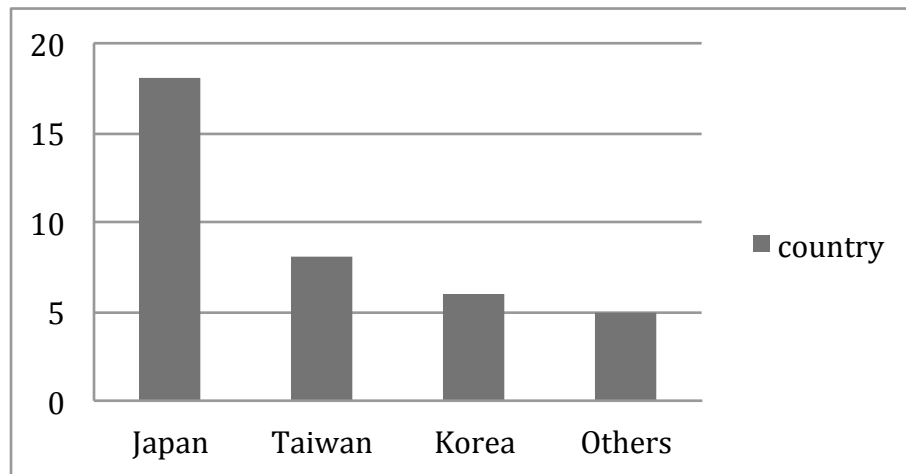
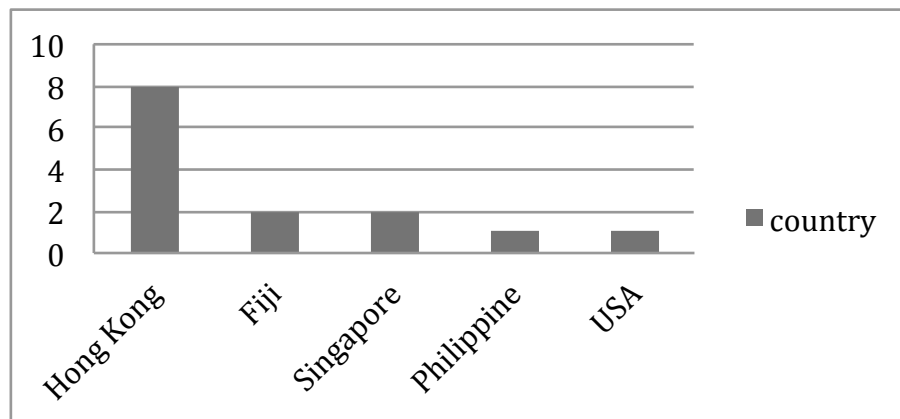**Figure 3**. Distribution of EFL settings.

Figure 4. Distribution of ESL settings.

(c) Age**.** A significant difference among different age groups was found ($Q = 9.14$, $p < 0.05$), which indicated that the age of the participants can influence the efficacy of the ER approach. The highest mean effect size was found in adults group ($d = 0.70$) followed by children ($d = 0.52$) and adolescents ($d = 0.35$) groups. The adult group, in particular, showed over twice as high an effect as the adolescent group. However, studies were heavily concentrated on adult groups ($k = 30$) which could have influenced the standard error of its effect. Currently, only six samples and 15 samples

were gathered for children and adolescent group, respectively. More research on these groups is needed to confirm this finding.

(d) Library size. The test of homogeneity indicated no statistical difference in library size. Therefore, the library size was found to have no significant influence on the impact of ER in this research.

(e) Length of treatment. No statistical difference was found among the length of programs. In fact, they showed very similar aggregated effect sizes (short: $d = 0.51$, medium: $d = 0.59$, long: $d = 0.60$) which implies that ER can be effective regardless of the program length. This was somewhat contradictory to the natural expectation that the long program would have greater impact on ER than short programs. The reason for this could be due to the fact that the long program had, much smaller sample size ($k = 4$) than the other two programs (*short*: $k = 16$, medium: $k = 31$).

(f) Text type. A significant difference between text types was found ($Q = 18.48$, $p < 0.05$). Surprisingly, web text ($d = 0.89$) showed almost double the effect of paper text ($d = 0.47$). This result is in accordance with Hung (2011)'s study which revealed that the ER of multimodal text was as effective as that of printed linear text in improving Taiwanese undergraduate EFL learners' English proficiency in simulated TOEIC tests. However, this result should be interpreted with caution since the web text group had smaller samples ($k = 10$) than the paper text groups ($k = 41$). Also, web text group was mostly comprised of adult groups which showed the highest aggregated effect size compared to other age groups. Only one study (i.e., Cho & Kim, 2004) had a children group.

(g) ER form. Type B (ER as a part of reading course) is found to be by far the most common type of ER form in the field (k = 38). The Q-test indicated a significant difference among variables ($Q = 30.90$, $p < 0.05$). ER as part of a curriculum showed the highest mean effect ($d = 0.91$), followed by ER as an extracurricular activity ($d = 0.67$), ER as a part of reading course ($d = 0.47$), and ER as an independent reading course ($d = 0.24$). However, the result should be interpreted with caution since ER forms are heavily concentrated in either Type B (ER as a part of reading course) or Type C (ER as a part of curriculum). Moreover, the studies in Type A (ER as an independent reading course) or Type D (ER as an extracurricular activity) were very hard to find. Only one study was found for each Type. Although Type A, which employs ER as an independent reading course, seems to be the most ideal way of implementing ER, it showed the smallest effect ($d = 0.24$). ER showed a higher effect when it was accompanied by other classroom activities such as interactive vocabulary instruction, group discussion, or oral presentation rather than devoting the entire class to only reading.

(h) Focus skills. No statistical difference was found among outcome variables, even though reading rate showed notably higher effect size ($d = 0.83$) than reading comprehension ($d = 0.54$) or vocabulary ($d = 0.47$). This may be due to the fact that the sample size of reading rate ($k = 10$) was not large enough to make a significant difference among variables.

Discussion

The first research question concerned the overall effectiveness of ER on reading proficiency (reading comprehension, reading rate, and vocabulary). The results showed that there was small to medium effect for both experimental-versus control group design ($d = 0.57$) and pre-to-post-test ($d = 0.79$) design. The effect can also be compared with regard to other meta-analytical findings in the field of second language studies. For example, the summary effect of corrective feedback was $d = 1.16$ (Russell & Spada, 2006), the effect of computer-assisted vocabulary instruction was $d = 0.75$ (Chiu, 2013), the effect of strategy instruction was $d = 0.49$ (Plonsky, 2011), and the effect of visual input enhancement was $d = 0.22$ (Lee & Huang, 2008).

Currently, however, an intensive reading approach is the major approach to teaching reading in the field and it is of no use if ER is not practiced in the school settings despite its reported benefits. One possible way to encourage the widespread use of ER is to inform and educate teachers, administrators, and policy makers of the effectiveness of ER. Without convincing them of its advantages over traditional teaching, it would be very difficult to adopt an ER approach in school settings.

First of all, policy makers and teachers should realize the fact that it takes time to see the benefits of ER (Lee, 2007). Therefore, once ER is implemented in the curriculum, it should not be removed quickly because its impact is not immediate or apparent. Also, teachers should change their view of considering ER as additional work that just takes up classroom time (Robb & Kano, 2013).

The second research question was whether the effects of ER differed significantly depending on the moderator variables. As for the identification variable, the number of studies on ER and the mean effect of ER increased gradually over last 30 years. This shows not only growing interest in ER but also developing expertise in ER classroom implementation, which suggests a promising perspective for its future.

As for contextual variable, settings and age seemed to affect the outcome of ER. Although an ESL setting is thought to be a better surrounding for learning English, an EFL setting can also produce positive effects despite its limited input environment. Moreover, an ER approach showed highest effect with adults and lowest effect with adolescents. The high mean effect of adults was also reported in Nakanishi (2015) and low mean effect of junior high students and high school students were also found in Kim (2012). This may be due to the fact that the adult group is better equipped cognitively than other age groups to start reading extensively. In other words, adults have better capacity to take advantage of ER since they have more experience, background knowledge, and vocabulary. The lowest impact of ER on adolescents, on the other hand, could be due to the test-centered curriculum in the schools. Despite the fact that adolescents are superior in their cognitive ability compared to children, they may not have any interest in ER since it does not have a direct impact on their grades.

One of the biggest hurdles of implementing ER is setting up a library, which, in fact, can require

quite a large budget. In this study, the library size was found to have no significant influence on the impact of ER. Hence, teachers can start ER on a small scale in their classrooms when the school cannot provide a library. Moreover, the high effect of web-based stories suggests the possibility of adopting a computer reader program instead of setting up an expensive library. Hung (2011), to illustrate, reported that the higher effect of multimodal text (i.e., computer reader program) was found in comparison with linear text (i.e., paper books) for EFL undergraduates. Although most of the studies included in the meta-analysis had a class or a school library with a wide range of books where students could freely borrow the books of their choice, many schools, in reality, may not have the funding necessary to set up a library.

A computer reader program can be a great substitute for a library. Moodle Reader (Robb & Kano, 2013), for example, not only provides students with vast amount of reading material, but it can also track students' reading progress and comprehension. It can be used outside of the classroom, so it does not intrude on classroom time. Moreover, it saves teachers from a lot of work such as going over students' reports to confirm their understanding. Furthermore, students' work on Moodle Reader is maintained in Excel documents which can be incorporated into a student's final grade. In short, a computer reader program is a cost-effective way to make ER more accessible.

However, it should be born in mind that which text type (paper books or digital books) is better for promoting reading proficiency is still controversial in both English as a first language and English in second or foreign language settings. For example, Empey (2013) reported negative evidence for digital books, claiming that it is difficult to keep focus on digital texts. Pino-Silva (2006) and Hung (2011), on the other hand, reported positive evidence such as enjoying the benefit of having access to hundreds of up-to-date magazines through the web.

Moreover, some students, particularly young learners, may not feel comfortable about using computers for reading or may not be ready to read independently outside of the classroom. In such cases, it may be better to set up a small classroom library. Teachers may prefer to set up a library with classical literature; but the readers' interests come first because students will not read extensively or voluntarily unless they are interested in the material. Teachers can find out their students' interests through questionnaires and purchase books or magazines accordingly (Day & Bamford, 1998).

Finally, the lack of class time issue in ER can be resolved by implementing ER as a part of the school curriculum. In this way, students' reading is not limited to classroom hours but is extended to reading outside the classroom. Whereas other ER forms engage students' interest in reading for a short term (i.e., only while research is conducted), incorporating ER into the curriculum can motivate students to read more over time. Moreover, implementing ER as a part of the school curriculum showed the highest mean effect among four different ER types.

Conclusion

This study examined the overall effectiveness of ER on reading proficiency in ESL and EFL settings. It also investigated how identification, context, treatment, and outcome variables

influenced the outcome of ER programs. Our meta-analysis demonstrated the overall effectiveness of an ER approach compared to an intensive or traditional reading approach. Also, the effect of ER can be maximized with the adults group, computer reader programs, and ER in the curriculum. Finally, the results showed a promising outlook for practicing ER in EFL settings.

Due to the financial and time issues, however, ER is yet not widely practiced in the field, particularly in EFL settings. Nevertheless, this study showed ER implementation can be made easy when ER becomes a part of a school curriculum and the cost of setting up a library can be significantly reduced by replacing paper books with digital books. Nowadays, technologies are evolving rapidly and the literacy of multimodal text is indispensable in this era of Information Computer Technology (Kress, 2003).

Although an individual teacher can practice ER program in his or her own classroom, ER programs can be greatly facilitated by systematic support from the schools or the government. For example, the schools or the government can provide a variety of books through the school library or computer programs which an individual teacher cannot afford on his or her own.

We hope this study provides a helpful and insightful summary of empirical findings on an ER approach in EFL and ESL settings. A limitation of this study is that most of the studies included in the analysis were from an Asian Pacific context, which may not reflect a balanced view of ER practice around the world. Also, long-term studies that go over a year and children-participated studies were very hard to find. Consequently, this resulted in a small number of samples in long-term studies and children-participated studies which may not have been large enough to reflect the true effect of ER in the long run and on children. More research on these topics is needed.

References

Adams, M. (1994). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

*Alavi, S., & Keyvanshekouth, A. (2012). Using the MoodleReader as an extensive reading tool and its effect on Iranian EFL students' incidental vocabulary learning. *English Language Teaching*, *5*, 135–145.

*Al-Homoud, F., & Schimitt, N. (2009). Extensive reading in a challenging environment: a comparison of extensive reading and intensive reading approach in Saudi Arabia. *Language Teaching Research*, *13*, 383–401. DOI:10.1177/1362168809341508

Bamford, J., & Day, R.R. (Eds.). (2004). *Extensive reading activities for teaching language.* New York: Cambridge University Press.

*Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese University EFL learners' reading rates. *Language Learning*, *62*, 665–703. DOI: 10.1111/j.1467-9922.2011.00651.x

*Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix, 1*, 27–45.

*Burrows, L. (2012). The effects of extensive reading and reading strategies on reading self-efficacy. Unpublished doctoral dissertation, Temple University. Tokyo, Japan.

*Cha, J.-E. (2009).The effects of extensive reading on enhancing vocational high school students' L2 vocabulary & reading rates. *English Teaching*, *64*, 3–30.

*Chen, C.-N., Chen, S.-C., Chen, S.-H., & Wey, S.-C. (2013). The effects of extensive reading via e-books on tertiary level EFL students' reading attitude, reading comprehension and vocabulary. *The Turkish Journal of Educational Technology*, *12*, 303–312.

Chiu, Y.-H. (2013). Computer-assisted second language vocabulary instruction: A meta-analysis. *British Journal of Educational Technology*, *44*, E52–E56. DOI: 10.1111/j.1467-8535.2012.01342.x

*Cho, K.-S., & Choi, D,-S. (2008). Are read-alouds and free reading "natural partners?" An experimental study. *Knowledge Quest*, *36*, 69–73.

*Cho, K.-S., Kim, H.-J., & Krashen, S. (2004). Recreational reading in English as a foreign language in Korea: positive effects of a sixteen-week program. *Knowledge Quest*, *32*, 36–38.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York, NY: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

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

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

*de Morgado, N. F. (2009). Extensive reading: Students' performance and perception. *The Reading Matrix*, *9*, 31–43.

*Elley, W. B. (1991) Acquiring literacy in a second language: The effect of book-based programs. *Language Learning*, *41*, 375–411. DOI: 10.1111/j.1467-1770.1991.tb00611.x

*Elley, W. B., & Mangubhai, E. (1983). The impact of reading on second language learning. *Reading Research Quarterly*, *19*, 53–67.

Empey, W. (2013). Digital vs. paper-based extensive reading: Affect and attitudes in Korean EFL learners. Unpublished master thesis, Daegu Haany University.

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

Greenberg, D., Rodrigo, V., Berry, A., Brinck, T., & Joseph, H. (2006). Implementation of an extensive reading program with adult learners. *Adult Basic Education*, *16*, 81–97.

*Hafiz, F. M., & Tudor, I. (1989). Extensive reading and the development of language skills. *ELT Journal*, *43*, 4–13. DOI: 10.1093/elt/43.1.4

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

Hedges, L. S. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*, 107–128.

*Horst, M. (2005). Learning L2 vocabulary through extensive reading: a measurement study. *The Canadian Modern Language Review*, *61*, 355–382.

*Huang, H.-T., & Liou, H.-C. (2007). Vocabulary learning in an automated graded reading program. *Language Learning & Technology*, *11*/3, 64–82.

Hung, S.-S. (2011) Extensive reading for undergraduate EFL learners: multimodal text vs. linear text. Unpublished doctoral dissertation, Washington State University.

Im, B.-B., Ahn, H.-S., & Yoon, S.-D. (2010). The effect of extensive reading on improving middle school students' writing ability. *Journal of the Korea English Education Society*,

9, 227–249.

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

*Jeon, J. (2008). Extensive reading in a formal English reading class. *English Teaching*, *63*, 49–83.

Kim, H.-Y. (2002). A study on realities of English reading instruction in Korean high schools. Unpublished master's thesis, Chonnam National University.

Kim, J. (2012). A meta-analysis of extensive English reading researches. *English Language & Literature Teaching*, *18*, 85–106.

*Kim, J. (2014). Extensive reading of Korean university EFL learners: Reading ability and lexical profiles in L2 writing. Unpublished doctoral dissertation, Hanyang University.

*Kim, J., & Hwang, J. (2006). A study of running extra-curricular classes in an elementary school using extensive reading program. *English Language Teaching*, *18*, 243–270.

*Kim, K., & Lim, M (2012). Improving English reading competence through extensive reading. *English*, *25*, 155–168.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for input hypothesis. *The Modern Language Journal*, *73*, 440–464. DOI: 10.2307/326879

Krashen, S. (2007). Extensive reading in English as a foreign language by adolescents and young adults: a meta-analysis. *The International Journal of Foreign Language Teaching*, *3*, 23–29.

Kress, G. (2003). *Literacy in the new media age.* London: Routledge.

Kweon, S.-O., & Kim, H.-R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language*, *20*, 191–215.

*Lai, .F-K. (1993a). Effect of extensive reading on English Language learning in Hong Kong. *CUHK (Chinese University of Hong Kong) Education Journal*, *21*, 23–36.

*Lai, F.-K. (1993b). The effect of a summer reading course on reading and writing skills. *System*, *21*, 87–100. DOI: 10.1016/0346-251X(93)90009-6

Lao, C. Y., & Krashen, S. (2000). The impact of popular literature study on literacy development in EFL: More evidence for the power of reading. *System*, *28*, 261–270.

Lee, S.-K., & Huang, H.-T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, *30*, 307–331. DOI: http://dx.doi.org/10.1017/S0272263108080479

*Lee, S.-Y. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal*, *38*, 150–170.

Lin, L.-F. (2010). Senior high school students' reading comprehension of graded readers. *Journal of Language Teaching and Research*, *1*, 20–28.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications, Inc.

*Lituanas, P., Jacobs, G. M., & Renandya, W. A. (2001). An investigation of extensive reading with remedial students in a Philippines secondary school. *International Journal of Educational Research*, *35*, 217–225.

*Mason, B. (2003). A study of extensive reading and the development of grammatical accuracy by Japanese university students learning English. Unpublished doctoral dissertation. Temple University, Tokyo.

*Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, *25*, 91–102. DOI: 10.1016/S0346-251X(96)00063-2

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

*Mermelstein, A. D. (2014). Improving EFL learners' reading levels through extensive reading. *The Reading Matrix*, *14*, 227–242.

Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly*, *49*, 6–37. DOI: 10.1002/tesq.157

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

*Park, S.-S., & Kang, W.-H. (2004). A study on the application of extensive reading of language learner literature in secondary English education. *Journal of the Korean English Education Society*, *3*, 97–138.

Pino-Silva, J. (2006). Extensive reading through the Internet: Is it worth the while? *The Reading Matrix*, *6*, 85–96.

Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, *61*, 993–1038. DOI: 10.1111/j.1467-9922.2011.00663.x

Plonsky, L., & Oswald. F. L. (2014). How big is "Big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. DOI: 10.1111/lang.12079

*Rezaee, A. A., & Nourzadeh, S. (2011). Does extensive reading improve EFL learners' processing ability? *Theory and Practice in Language Studies*, *1*, 1167–1175.

*Robb, T. N., & 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.

Rodrigo, V., Krashen, S., & Gribbons, B. (2004). The effectiveness of two comprehensible-input approaches to foreign language instruction at the intermediate level. *System*, *32*, 53–60. DOI:10.1016/j.system.2003.08.003

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Philadelphia: John Benjamins.

*Sheu, S. (2003). Extensive reading with EFL learners at beginning level. *TESL Reporter*, *36*, 8–26.

*Shin, K.-C. (2013). Vocabulary acquisition through extensive reading in EFL class. *Modern English Education*, *4*, 3–16.

*Sims, J. M. (1996). A comparative study of improvements in reading comprehension of skill-based instruction and extensive reading for pleasure with Taiwanese freshmen university students. Unpublished doctoral dissertation, Florida State University.

*Smith, K. (2006). A comparison of "pure" extensive reading with intensive reading and extensive reading with supplementary activities. *International Journal of Foreign Language Teaching*, *2*, 12–15.

*Taguchi, E., Takayasu-Maass, M., & Gorsuch, G. J. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language*, *16*, 70–96.

*Takase, A. (2007). Japanese high school students' motivation for extensive L2 reading. *Reading in a Foreign Language*, *19*, 1–18.

*Takase, A. (2009). The effects of SSR on learners' reading attitudes, motivation and
    achievement: A quantitative study. In A. Cirocki (Ed.), *Extensive reading in English
    language teaching* (pp. 547–559). Munich, Germany: Lincom.
*Tanaka, H., & Stapleton, P. (2007). Increasing reading input in Japanese high school EFL
    classrooms: An empirical study exploring the efficacy of extensive reading. *The Reading
    Matrix*, *7*, 115–131.
*Tsang, W.-K. (1996). Comparing the effects of reading and writing on writing performance.
    Applied Linguistics, *17*, 210–233. DOI: 10.1093/applin/17.2.210
*Weitz, W. E. (2003). Sustained silent reading with non-native speakers of English: its impact on
    reading comprehension, reading attitude, and language acquisition. Unpublished doctoral
    dissertation, University of Southern California.
*Yamamoto, Y. (2011). Bridging the gap between receptive and productive vocabulary size
    through extensive reading. *The Reading Matrix*, *11*, 226–242.
*Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency.
    System, *36*, 661–672. DOI:10.1016/j.system.2008.04.003
*Yang, E.-M. (2010). A study on the effectiveness of extensive reading in college English
    classes. *English Language & Literature Teaching*, *16*, 169–188.
*Zimmerman, C. B. (1997). Do reading and interactive vocabulary instruction make a difference?
    An empirical study. *TESOL Quarterly*, *31*, 121–140. DOI: 10.2307/3587978

* indicates studies meta-analyzed in this research.

About the Authors

Eun-Young Jeon is a lecturer at Chonbuk National University. She was a research scholar at
University of Hawai'i from 2013 to 2015. Her research interests include extensive reading in
English, ELT methodology, and material development. E-mail: joyjeon@jbnu.ac.kr

Richard R. Day is a Professor in the Department of Second Language Studies, University of
Hawai'i. He is the author of numerous publications, particularly on second language reading. E-
mail: rday@hawaii.edu