The Impacts Of Extensive Reading on English Vocabulary Learning: A Meta-Analysis

M. Farkhi Faishol Hakim, mfarkhi.20013@mhs.unesa.ac.id, UNESA, Indonesia
Pratiwi Retnaningdyah, tiwik.pr@gmail.com, UNESA, Indonesia
Syafi'ul Anam, syafiulanam@unesa.ac.id, UNESA, Indonesia

Abstract

Extensive reading has been continuously studied as a promising instructional method for improving students' language proficiency, including reading proficiency, vocabulary acquisition, and grammar awareness. The present study is a meta-analysis, which synthesized the data of 21 empirical studies (N = 1268). It was designed to explore whether extensive reading instruction effectively improved students' vocabulary acquisition and how the effectiveness varied in terms of the instruction length and teaching methods. Stata 14.0 was utilized to calculate the collected data. The results revealed that: (1) extensive reading has a significant effect on English vocabulary learning; (2) one semester (less than three months) is the most appropriate length of extensive reading instruction for vocabulary learning; (3) Graded Readers, comprehension questions and vocabulary exercise play significant roles as reading materials and education methods in promoting the vocabulary learning of EFL learners.

Keywords: a meta-analysis, extensive reading, vocabulary, effectiveness, EFL

1. Introduction

An increasing number of empirical studies have been conducted to examine the effectiveness of extensive reading all over the world in the past decades (e.g., Lee, 2007; Kweon & Kim, 2008; Cha, 2009; Yamashita & Kan, 2010; Soltani, 2011; Tiryaki & Tütüniş, 2012; Chang, 2013; Hamed, 2014; Webb & Anna, 2015; Ismael, 2017; Suk, 2017; Tabata-Sandom, 2017). However, the research limitation of each independent study, such as the relatively short duration of the treatment, small sample size and the lack of replication, may cause difficulty for researchers and English teachers to determine whether extensive reading is universally effective. Furthermore, how it affects the English vocabulary acquisition specifically and how reliable the outcomes of educational treatments may cause similar difficulties. This study adopts a meta-analysis to investigate the overall effectiveness of extensive reading on vocabulary learning in an EFL environment. Besides examining the overall effectiveness of extensive reading on vocabulary learning, it also sheds light on the specific treatment in comprehensive reading instruction, including reading materials and teaching methods. In this study, researchers and teachers concerned with extensive reading may find new research directions and be inspired to design more effective comprehensive reading instruction.

According to Glass (1976), meta-analysis is "the statistical analysis of a large collection of analysis results from individual studies to integrate the findings" (p. 3). A meta-analysis can

confirm the effectiveness and provide greater insight into a treatment based on the large size and better quality of primary studies.

But that is not to say meta-analysis has no drawbacks. One of its main problems is publication bias. Publication bias happens when the published literature data are exclusively synthesized, unrepresentative of the population of complete studies (Rothstein et al., 2006). The meta-analysis should include data from both published and unpublished studies. It is usually required to do a sensitivity analysis to ascertain publication bias. Additionally, it is proposed that tight selection criteria include qualifying studies before aggregating papers of varying research quality.

Extensive Reading and Vocabulary Learning

Extensive Reading

Extensive reading is a type of reading instruction that provides many reading materials to learners independently. Students must read extensively for general meaning, information, and pleasure (Day, 2004). Through extensive reading, learners can develop good reading habits, build vocabulary knowledge andenhance positive reading attitudes. In particular, it is believed that learners may encounter more unknown words when reading independently, which brings opportunities to infer the unfamiliar words in specific contexts and thus learn their meanings (Krashen, 1989). Although the mechanism is commonly accepted as accurate, as Cobb (2007) states, it is still disputable whether extensive reading is essential. So far, studies concerning extensive reading vary in terms of different study focus, including reading attitude (e.g., Chin–Neng, 2013), reading comprehension (e.g., Wijaya, 2013; Rezaee, 2011; Lin, 2010), reading speed (e.g., Al–Homoud, 2012; Beglar, 2012), grammar (e.g., Rodrigo, 2004) and vocabulary acquisition (e.g., Tan, 2016; Waring, 2003) etc.

Compared with extensive reading, intensive reading requires students to read short texts, obtain detailed meaning with close guidance from the teacher to acquire vocabulary and grammar knowledge and develop some reading skills such as identifying main ideas and recognizing text connectors (Hong, 2014). Palmer defines it as "take a text, study it line by line, referring at every moment to our dictionary and our grammar, comparing, analyzing, translating, and retaining every expression that it contains" (1964, p. 111).

Vocabulary Learning

Vocabulary refers to the knowledge of words and word meanings. It is suggested that "teaching vocabulary will not guarantee success in reading, just as learning to read words will not guarantee success. However, lacking either adequate word identification skills or adequate vocabulary will ensure failure" (Biemiller, 2005). According to the National Reading Panel's synthesis of vocabulary research, vocabulary instructions were suggestedto:

- be direct for a specific text (Nagy & Judith, 2000).
- increase the exposure to vocabulary items repeatedly (Stahl, 2005).
- cover vocabulary words that the learners will find helpful in many contexts.
- restructure vocabulary tasks as necessary.
- entails active engagement, which goes beyond definitional knowledge, so vocabulary learning iseffective.
- use computer technology effectively to help teach vocabulary.

· cover multiple vocabulary instruction methods, resulting in optimal learning (NICHD, 2000).

Previous Practice of Meta-Analysis on the Effectiveness of Extensive Reading

In recent years, meta-analysis has been utilized more frequently than ever before in second language acquisition. Nakanishi (2015) conducted a meta-analysis to examine the overall effectiveness of extensive reading on comprehensive language proficiency. However, few meta-analysis studies exclusively focused on examining extensive reading's effects on vocabulary acquisition to the authors' knowledge. In consideration of the shortcomings of empirical studies and insufficiency of meta-analysis concerning the extensive reading's effects on English vocabulary acquisition, the primary purpose of this study is to synthesize the findings of studies on the extensive reading's effects on English vocabulary learning to draw more reliable conclusions regarding extensive reading's overall strength and the identification of the influencing variables.

In addition, the significant difference between this study and Nakanishi's (2015) lies in the study interest. Nakanishi studied the overall effect of extensive reading on different language proficiencies, including reading speed, reading comprehension, vocabulary and grammar. In contrast, the present study exclusively focuses on specific aspects of language proficiency and vocabulary learning and conducts detailed exploration of the particular treatment in the independent studies.

Research Questions

1) Does extensive reading positively impact the English vocabulary acquisition of EFL learners?

2) What is the best length of treatment in extensive reading programs?

3) What are the most effective reading materials and pedagogical treatment of extensive reading in English language instruction?

2. Methodology

2.1 Data Selection and Inclusion Criteria

The following databases were used to search articles examined in this current meta-analysis: *Google Scholar, the Education Resources Information Center (ERIC), ProQuest* and *Cambridge Core*. As a result, our data were composed of published works of the following 16 journals: *Advances in Language and Literary Studies, Applied Linguistics – Global and Local, Brazilian, English Language Teaching Journal (BELT), ELT Research Journal, English Language Teaching, English Teaching, International Journal of English and Education, Language Teaching Research, Malaysian Journal of ELT Research, Mediterranean Journal of Social Sciences, Reading in a Foreign Language, RELC Journal, Studies in Literature and Language, Studies in Second Language Acquisition, The Reading Matrix* and *Vocabulary Learning and Instruction.*

Unpublished studies (two doctoral dissertations) were also included because the publication bias will occur if allthe studies included in a meta-analysis are published.

Three criteria were employed to assess the collected studies from the abovementioned sources.

1) The studies should be empirical, which means experiments concerning the effects of ER on vocabularyacquisition should be conducted.

2) The studies should be published or conducted between 2007 and 2016.

3) Data contained in the studies should be available, e.g. means and standard deviations were presented.

With the criteria given above, 21 studies, asterisked in the References section, were finally included in this meta-analysis, in which 17 have control groups, and the remaining four have not.

2.2 Data Coding

To see the overall and specific effects of different variables, the authors coded all 21 studies. They categorized all the study variables into seven: participants, length of instruction, control group, reading materials, treatment, test use and test reliability (see Table 1).

Participants

This moderator will help people know which age group's vocabulary learning has benefited from extensive reading and how many studies have focused on that age group.

Length of Instruction

As a continuous variable in any extensive reading program, the length of instruction will impact the learners' level of vocabulary learning through extensive reading and answer the second research question mentioned before how long the best extensive reading treatment period will be. Length of instruction was divided into three dichotomous categories for the majority of the 21 studies.

Control Group

Given that some extensive reading empirical studies do not have control groups to compare the effects of the treatments, this moderator is crucial for this study to see whether the control group brings convenience to divide the overall impacts into two parts in an obvious way.

Reading Material

As one of the most engaging factors in extensive reading, reading material will answer the last research question of whether reading material is suitable for and widely used in comprehensive reading studies. As an online platform for students doing extensive reading and related exercises, MoodleReader provides texts from Graded Readers for students to read. Stories included in the chapter book are selected from various storybooks. In this case, MoodleReader and chapter books were categorized as Graded Readers and storybooks, respectively.

Treatment

As one of the creativities and significances of this meta-analysis, the current study will help readers find out which method in extensive reading will be the most effective and thus propose pedagogical implications for teachers. Five subfactors, book report, comprehension questions, dictionary usage, vocabulary exercise and vocabulary instruction, are included in

the treatment. Vocabulary exercise includes sentence making and vocabulary worksheet. The teachers or researchers designed them to strengthen

Students' vocabulary was learned after class. Vocabulary instruction refers to teachers' direct teaching of unknown oressential language in the extensive reading program. In addition, other treatments like MoodleReader and oral rendition of texts have also been reported to improve vocabulary acquisition. All the variables included in treatment were declared to be directly effective to students' vocabulary acquisition in the target 21 studies.

Test Use

Tests including pretests and posttests are exclusively vocabulary tests by which the mean scores and standard deviations were generated and collected. The version of the test used in the treatment influences the results of the empirical studies, so the reporting of the test version is essential.

Reliability

Whether the used tests are reliable is also essential for any study to report, which will affect the reliability of this meta-analysis.

Table 1. Data coding

Variables	Codes
Participants	1. Junior high school students
	2. High school students
	3. University students
	4. Adults
	5. Children (elementary school or below)
Length of instruction	1. One semester (less than three months)
	2. Two semesters [3 months to 6 months]
	3. One year (6 months, one year]
Reading materials	1. Graded Readers
	2. Storybooks
	3. Reading power
Treatment	1. Book report
	2. Comprehension questions
	3. Dictionary usage
	4. Vocabulary exercise
	5. Vocabulary instruction
	6. Others
Control group	1. None (i.e., pre-post design)

	2. One group
	3. Two groups or more
Test use	1. Same test version used in pre and posttest
	2. Parallel version
Test reliability	1. Not reported
	2. Reported (data based on the current study)
	3. Reported (data cited from the test manual)

2.3 Calculation and Interpretation of Effect Sizes

Among the 21 original studies, 17 have at least one control group, while the other four only involve experimentalgroups. Because most of the studies compared two groups, Cohen's *d* index was used as the standardized effect

Effect size. The descriptors for magnitudes of d–type effect size are displayed in Table 2 below. The *d* value from 0.01 to 2.0 was initially proposed by Cohen (1988) and later expanded by Sawilowsky (2009). In terms of the interpretation of effect size, the bigger the *d* value is, the more parallel in size the effect is (Ellis, 2010).

Table 2. The interpretation of effect sizes

Effect size	***d***	**Reference**
Very small	0.01	Sawilowsky, 2009
Small	0.2	Cohen, 1988
Medium	0.5	Cohen, 1988
Large	0.8	Cohen, 1988
Very large	1.2	Sawilowsky, 2009
Huge	2	Sawilowsky, 2009

The formulas illustrated in Appendix were applied to calculate Cohen's *d* with descriptive data given by the original studies. According to Ellis (2010), different studies could be compared with the effect size. Two formulas were used to calculate effect sizes for studies with and without control groups. The ultimate meansand standard deviations used in this meta-analysis were average values calculated by the authors.

2.4 Effect Model

Two statistical models can be adopted when performing a meta-analysis: fixed effect and random effect. The fixed–effect model is not concerned with dispersion in the observed effects because it is assumed to reflect nothing more than sampling error. In the random-effect model, the distribution in impact is considered authentic. In most cases, the usage of the model is determined according to the test result of heterogeneity: if $p > 0.1$, it is suggested to employ the fixed–effect model since homogeneity is observed among the independent studies; if $p \leq 0.1$, it is recommended to employ the random–effect model (Zhang, 2015).

Since the heterogeneity was observed in the present study ($p \leq 0.1$), the random effect model was used.

2.5 Research Synthesis

Cohen's *d* values were identified by calculating two contrastive means of dependent variables, between-group contrasts and within-group contrasts (pre-post Contrasts). Between-group contrasts involve 17 studies, including both experimental and control groups. Other four studies without control groups were examined through pre-post differences. Tables 3 and 4 show the aggregated effect sizes for group contrasts and pre–post contrasts, respectively. Twenty–one unique effect sizes (17 effect sizes for group contrasts and four effect sizes for pre-post contrasts) and a total sample size of 1,268 participants were included. The sample sizes of group contrasts ranged from 10 to 67 (M = 31.88) for experimental groups and 10 to 139 (M = 35.65) for control groups. The sample sizes of the pre-post contrasts ranged from 12 to 60 (M = 30).

After all, the descriptive statistics and the effect sizes were calculated, the data were then submitted to Stata (Version 14) to do a meta-analysis.

Table 3. Aggregated effect sizes (group contrasts)

	N (total)	*N exp*	*N c*	*d*
Alavi & Keyvanshekouh (2012)	38	20	18	1.25*
Al–Homoud & Schmitt (2009)	65	45	20	0.52*
Cha (2009)	20	10	10	0.45
Chang (2013)	64	30	34	-1.77
Hamed (2014)	36	18	18	4.56
Jafarpour (2014)	40	20	20	1.02
Lee (2007) Study 1	141	65	76	-0.77
Lee (2007) Study 2	206	67	139	0.17
Rafatbakhsh & Alavi (2013)	33	15	18	-0.13
Rashtchi & Pourmand (2014)	60	30	30	0.58
Rosszell (2007)	40	20	20	1.06*
Soltani (2011)	80	40	40	1.45*
Teng (2015) Study 1	52	26	26	4.77*
Teng (2015) Study 2	46	23	23	2.58*
Tiryaki & Tütüniş (2012)	100	50	50	5.18*
Vaezi & Nilforooshan (2013)	60	30	30	1.96*
Yamamoto (2011)	67	33	34	0.59*

Note. exp = experiment group. *c* = control group.

*Indicates an averaged effect size within the study.

Table 4. Aggregated effect sizes (pre-post contrasts)

	N exp	*d*
Chang (2015)	31	4.21*
Kweon & Kim (2008)	12	1.68*
Webb & Chang (2015)	60	1.24*
Yamashita & Kan (2010)	17	0.67*

Note. exp = experiment group.
*Indicates an averaged effect size within the study.

5. Results and Discussion

5.1 Features of the Data

Table 5 demonstrates the research features of studies covered in this meta-analysis and the corresponding percentages. The majority of the participants were university students (67%), followed by high school students (19%).

Table 5. Research features

Features		No. of studies	%
Participants	1. Junior high school students	1	5
	2. High school students	4	19
	3. University students	14	67
	4. Adults	1	5
	5. Children (elementary school or below)	1	5
Length of instruction	1. One semester (less than 3 months)	11	52
	2. Two semesters (from 3 months to 6 months)	8	38
	3. One year (from more than 6 months to 1 year)	2	10
Reading materials	1. Graded Readers	18	86
	2. Story books	6	29
	3. Reading Power	2	10
Treatment	1. Book report	15	71
	2. Comprehension questions	3	14
	3. Dictionary usage	7	33
	4. Vocabulary exercise	8	38
	5. Vocabulary instruction	4	19
	6. Others	6	29
Control group	1. None (i.e., pre–post design)	4	19

	2. One group	15	71
	3. Two groups or more	2	10
Test use	1. Same test version used in pre and post test	12	57
	2. Parallel version	9	43
Test reliability	1. Not reported	12	57
	2. Reported (data based on the current study)	5	24
	3. Reported (data cited from the test manual)	4	19

Note. k = 21. All the percentages are rounded. The total number of studies adopting reading materials and treatment varies because some studies included more than one reading material or treatment type.

More than half of the studies had short–term instruction (less than three months), and 38% of the studies' instructions were between 3 months and six months. Only two studies were conducted between 6 months and one year. However, studies conducted for more than one year were not included.

In terms of reading materials, Graded Readers (86%) was most used among the 21 studies, whereas storybooks and Reading Power were only used in 6 and 2 studies, respectively. As regards extensive reading treatment concerning students' vocabulary acquisition, book report (71%) was the top option, followed by vocabulary exercise (38%), dictionary usage (33%) and others (29%), including MoodleReader and oral rendition of texts. On the contrary, vocabulary instruction (19%) and comprehension questions (14%) were the least popular. Several 81% had one or more control groups, while 19% failed.

Over half of the studies administered pre–posttests whose versions were the same, and 43% of the studies used parallel versions of tests. Only 24% of the studies report test reliability based on the current research, but 19% cited data from the tests manual.

5.2 Publication Bias (Sensitivity Analysis)

It is essential to assess publication bias in the meta-analysis because "literature reviews regarding support for a hypothesis can be biased if the original literature is contaminated by publication bias" (Rothstein et al., 2006). The most common way to present publication bias is the employment of a funnel plot which indicates a relationship between effect size and study precision. To measure study precision, choices including sample size, standard error, and inverse effect size variance are available. However, Sterne and Egger (2001) conclude that common error is the most recommendable based on their comparison.

As Ioannidis (2007) claims, a funnel plot can be drawn when at least ten studies are in the meta-analysis. Therefore, the funnel plot for pre–post-contrast fails to be provided because the number of studies in pre-post contrasts is only four. Figure 1 shows the funnel plot for group contrasts, visually representing the possible publication bias. A relationship between standard errors plotted on the reversed y–axis and effect sizes (SMD) for each study plotted on the x–axis is displayed in the funnel plot.

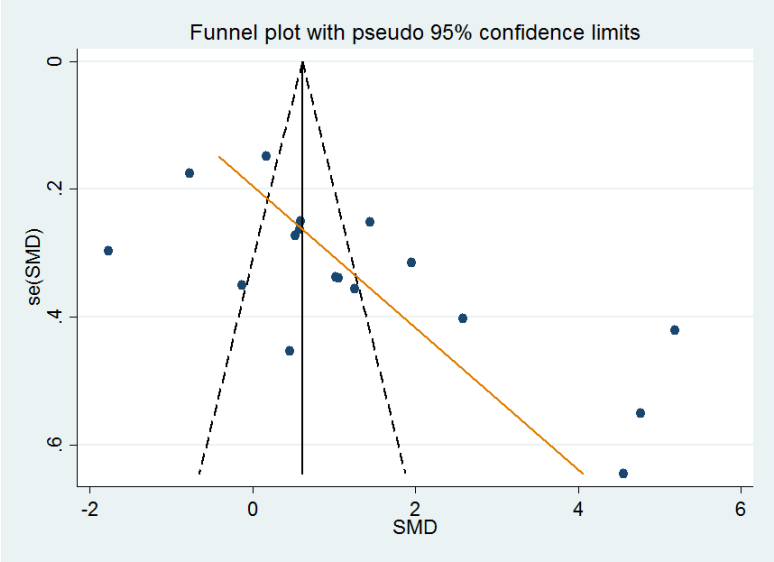

Figure 1. Funnel plot of group contrasts

It should be noted that the most extensive studies have minor standard errors, which is the reason why the most comprehensive studies are placed at the top of the graph; thus, the y–axis must be reversed (standard error 0 at the top) (Sterne et al., 2004).

Because the accuracy of the effect will increase when the number of studies is significant, smaller effects will be scattered widely at the bottom of the plot, and more significant effects will be distributed intensively in the centre or top of the funnel plot. In addition, among many reasons that may cause the asymmetry of the funnel plot, the major one might be the uneven quality of the studies. Smaller and low-quality studies may trigger significant intervention because of their less precise data.

In Figure 1, the funnel plot presents an asymmetric inverted funnel, and the distribution is relatively dispersed. Therefore, the included studies may have publication bias. The reason might be that the research sizes included in this meta-analysis are relatively small. Therefore, more large-size studies should have been included in this meta-analysis. In addition, the data adopted in this meta-analysis is continuous, and standardized mean difference (SMD) is used. There is no recommendable method to exam the asymmetry of the funnel plot because the number of studies using continuous data and the standardized mean difference is not big enough for reference (Zhang, 2015). Thus, the asymmetry of the funnel plot will not be further examined.

5.3 Meta-Analysis

To answer the first research question, the authors used Cohen's *d* value to examine the overall effect size of extensive reading's effects on English vocabulary acquisition. The results of the group contrasts are illustrated in Table 6, including effect sizes, the number of studies, sample size, and confidence intervals.

Table 6. Overall meta-analysis results for group contrasts

Group	Subgroup	*d*	*k*	*N*	*CI* (Low, High)
Overall		1.32	17	1,148	[0.62, 2.02]
Participants	1. Junior high school students	0.58	1	60	[0.06, 1.09]
	2. High school students	0.82	2	60	[0.28, 1.35]
	3. University students	1.1	12	868	[0.33, 1.87]
	4. Adults	1.96	1	60	[1.34, 2.57]
	5. Children (elementary school or below)	5.18	1	100	[4.36, 6.01]
Length of instruction	1. One semester (less than three months)	2.1	9	607	[0.83, 3.37]
	2. Two semesters [3 months, 6 months]	0.56	7	335	[-0.37, 1.49]
	3. One year (6 months, one year]	0.17	1	206	[-0.12, 0.46]
Reading materials	1. Graded Readers	1.48	14	975	[0.63, 2.34]
	2. Story books	0.62	5	419	[0.06, 1.17]
	3. Reading power	0.5	2	85	[0.05, 0.96]
Treatment	1. Book report	0.57	11	787	[-0.07, 1.21]
	2. Comprehension questions	4.77	1	52	[3.69, 5.85]
	3. Dictionary usage	1.3	5	221	[0.35, 2.24]
	4. Vocabulary exercise	1.35	8	396	[0.55, 2.15]
	5. Vocabulary instruction	1.46	4	346	[-0.66, 3.58]
	6. Others	0.32	4	195	[-1.38, 2.03]
Test use	1. Same test version used in pre and post	1.45	9	735	[0.47, 2.43]
	2. Parallel version	1.18	8	413	[0.08, 2.29]
Test reliability	1. Not reported	0.71	8	682	[-0.38, 1.79]

	2. Reported (data based on the current study)	0.92	5	292	[0.42, 1.42]
	3. Reported (data cited from the test manual)	3.18	4	174	[1.41, 4.96]

As shown in Table 6, the overall confidence interval (CI) does not include 0, which means that extensive reading significantly affects English vocabulary learning (d = 1.32), and experimental groups perform better than the control groups. In terms of the interpretation of the test of homogeneity, Higgins (2003) suggested that the bigger I^2 is, the higher the homogeneity is. The result (p = 0.00, $d.f.$ = 16, I^2 = 96.1%) of the homogeneity test indicates that homogeneity exists and is quite large.

Furthermore, the overall effect size (d = 3.26) in Table 7 indicates extensive reading in the pre-post contrasts. Participants perform better in the posttests than in the pretests. This effect size is much larger than group contrasts (d = 1.32). As a result, (p = 0.008, $d.f.$ = 3, I^2 = 22.2%) of the homogeneity test suggests the homogeneity is slight and not statistically significant because I^2 between 0% and 40%.

Table 7. Overall meta-analysis results for pre-post contrasts

Group	Subgroup	*d*	*k*	*N*	*CI* (Low, High)
Overall		3.26	4	120	[1.74, 4.77]
Participants	1. Junior high school students	-	-	-	-
	2. High school students	3.85	2	91	[2.17, 5.53]
	3. University students	0.71	2	29	[-2.77, 4.18]
	4. Adults	-	-	-	-
	5. Children (elementary school or below)	-	-	-	-
Length of instruction	1. One semester (less than 3 months)	4.19	2	43	[2.41, 5.97]
	2. Two semesters [3 months, 6 months]	0.67	1	17	[-2.87, 4.21]
	3. One year (6 months, 1 year]	1.24	1	60	[-3.56, 6.04]
Reading materials	1. Graded Readers	3.26	4	120	[1.74, 4.77]
	2. Storybooks	1.68	1	12	[-16.68,20.04]
	3. Reading power	-	-	-	-
Treatment	1. Book report	3.26	4	120	[1.74, 4.77]
	2. Comprehension questions	4.19	2	43	[2.41, 5.97]
	3. Dictionary usage	3.85	2	91	[2.17, 5.53]

	4. Vocabulary exercise	-	-	-	-
	5. Vocabulary instruction	-	-	-	-
	6. Others	3.85	2	91	[2.17, 5.53]
Test use post	1. Same test version used in pre and	0.89	3	89	[-1.93, 3.71]
	2. Parallel version	4.21	1	31	[2.42, 6.01]

5.4 Moderator Variable Analysis

Research questions 2 and 3 are concerned with investigating systematic disparity in the effectiveness of extensive reading across age groups, length, reading materials, treatment and test of instruction. Effect sizes were calculated for coded subgroups (see Table 6 and Table 7).

First of all, five age group variables were investigated. Regarding group contrasts, all the confidence intervals do not include zero. The effect size for children indicates a huge effect (d = 5.18), followed by adults (d = 1.96), university students (d = 1.10), high school students (d = 0.82) and junior high school students (d = 0.58). For the pre-post contrasts, the effect size for university students includes zero. And there is a huge effect for high school students (d = 3.85).

In terms of the length of extensive reading instruction, effect sizes for two semesters and one year in the group contrasts include zero, while one semester (d = 2.10) has a significantly huge effect. Similarly, the confidence intervals for two semesters and one year in the pre-post contrasts include zero, while one semester (d = 4.19) has a considerable effect size. Overall, the shorter the length of extensive reading instruction, the larger the effect is.

Regarding reading materials in the group contrasts, a very large effect for Graded Readers (d = 1.48), and medium effects for storybooks (d = 0.62) and reading power (d = 0.50) are obtained. In the pre-post contrasts, the confidence interval for storybooks includes zero. A huge effect for Graded Readers (d = 3.26) is obtained. Graded Readers is the most appropriate reading material for extensive reading instruction.

Concerning treatment in the group contrasts, the three confidence intervals for a book report, vocabulary instruction and others include zero. Comprehension questions (d = 4.77) has huge effect size; dictionary usage (d = 1.30) and vocabulary exercise (d = 1.35) have very large effect sizes. Likewise, in the pre-post contrasts, stu*d*ies adopting book report (d = 3.26), comprehension questions (d = 4.19), dictionary usage (d = 3.85) and others (d = 3.85) produce huge effect sizes. Overall, comprehension questions are the most effective ER instruction method, followed by dictionary usage.

Furthermore, test use produces the following effects. In group contrasts, both studies using the same test version for pre–and posttest (d = 1.45) and studies using parallel version (d = 1.18) produce huge effect sizes. However, the confidence interval for studies using the same test version includes zero in pre-post contrasts. And the effect size of studies using parallel version (d = 4.21) is much huger than the one for group contrasts.

Regarding the test reliability, the confidence interval for studies without reported test reliability include zero. In contrast, studies reporting reliability with data based on the current research (d = 0.92) yield a significant effect, and studies reporting reliability with cited ones (d = 3.18) produce a huge effect size. Notice that all the four investigations in pre-post contrasts do not report test reliability, so it is meaningless to display the results.

The Vocabulary Levels Tests (VLT) (Schmitt et al., 2001) were used most frequently (n = 8), followed by the Vocabulary Levels Test (n = 3). Ten of the remaining studies used vocabulary tests developed by previous researchers, such as the Productive Vocabulary Levels Test (Laufer & Nation, 1999) and the Vocabulary Knowledge Scale (Wesche and Paribakht, 1996), while the remaining investigations used their general vocabulary knowledge assessments.

Overall Effect of Extensive Reading on Vocabulary Learning

The first research question is concerned with the effect of extensive reading instruction. In the group contrasts (d = 1.32), students who received extensive reading instruction outperformed students who did not respond to a large extent; similarly, in the pre-post differences (d = 3.26), students made huge progress significantly in vocabulary learning after receiving extensive reading instruction. It should be noted that these huge effect sizes are consistent with the results of studies reviewed in this synthesis, which soundly confirms that extensive reading has unquestionably a positive impact on students' English vocabulary acquisition to a huge extent. There is no adverse effect because the effect sizes for pre-post contrasts ranged from 1.74 to 4.77. However, it is suggested that its effect size (d = 3.26) should be interpreted cautiously since control groups were not included in the analysis.

5.5 Learner's Age

Although learners' age is not involved in the research questions, it is worthy of finding out which age group benefits the most from extensive reading instruction on their vocabulary learning and how they are different. Surprisingly, extensive reading was found to be effective in all group contrasts of different ages, and the effect sizes for children (d = 5.18) and adults (d = 1.96) show to be huge. In the pre-post contrasts, no effect was found among university students, while the effect size for high school students (d = 3.85) is quite huge, which might be caused by the small number of studies included. Two studies for each were included in terms of high school students and university students. Junior high school students, children and adults only have one study respectively. In the group contrasts, the finding that children benefit the most differs from that of Nakanishi (2015) because Nakanishi didn't include children participants in his study. However, we should note that the limitation of participants' study number (k = 1), and further research is recommended. Among the other age groups, adults (d = 1.10), university students (d = 0.82), and junior high school students (d = 0.58), all have a substantial effect size. It may indicate that the older they are, the more successful extensive reading promotes their vocabulary acquisition. This finding is in line with Nakanishi's study, which concludes that extensive reading might be more beneficial for late learners because of their more vital analytical skills in understanding the reading contents.

It should be noted that the numbers of studies for the participant categories are pretty small except university students. It is strongly suggested that more research should be conducted to

confirm this finding. There is no difference among groups as the overlapping confidence intervals indicate statistically.

5.6 Length of Instruction

The second research question is concerned with the length of extensive reading instruction. As the results indicate, one semester of comprehensive reading instruction for both group and pre-post contrasts yields huge effect sizes (d = 2.10 and d = 4.19 – respectively). Whereas no effect was found for two-semester and one–year instruction in both group contrasts and pre–post-contrast. Therefore, two exciting results were revealed. One aspect is that one semester seems to be the most appropriate instruction period. The other is that shorter-term ER instruction can promote vocabulary learning better, contrary to the finding of Nakanishi (2015), i.e., the longer the instruction length is, the better the treatment will be. This finding is also different from the claim of Grabe (2009, p. 328), "reading extensively when done consistently over a long period, leads to better reading comprehension as well as improved abilities in several other language areas". Based on Papagno's (1991) statement that phonological short–term memory contributes to foreign–language vocabulary learning, the reason for the results might be that a shorter period of time for extensive reading instruction is suitable for participants to memorize a certain amount of vocabulary. On the contrary, the longer length of education may lead to quick forgetting. However, this hypothesis should be proved further by more studies.

5.7 Reading Materials

The last research question is concerned with the procedure of extensive reading instruction. Specifically, this meta-analysis tries to determine the most effective reading material and the most effective education treatment in terms of vocabulary learning through extensive reading. In the group contrasts, three categories of reading materials are reported to be effective. Graded Readers (d = 1.48) produces a very large effect, followed by storybooks (d = 0.62) and reading power (d = 0.50) both yielding medium effect. Only Graded Readers (d = 3.26) produce a huge effect in the pre-post contrasts. The results indicate that Graded Readers is most suitable for participants to read among the three options for extensive reading instruction, which is consistent with the statement that Graded Readers are useful reading materials for students with a lower–intermediate level to improve the learning of vocabulary (e.g., Nation, 2001; Nation, 2009). Compared with Graded Readers, storybooks and the reading power are less effective, probably because the contents are not simplified, or the storybooks are not as strictly classified as Graded Readers. It is recommended that future studies should investigate the effect of storybooks and reading power on vocabulary learning in extensive reading instruction.

5.8 Treatment

Another crucial factor influencing the effect of extensive reading on vocabulary learning is the education treatment or method. Three categories, book report, vocabulary instruction and others, include the confidence interval of zero for group contrasts. At the same time, the effect size for comprehension questions (d = 4.77) produces a huge effect, followed by dictionary usage (d = 1.30) and vocabulary exercise (d = 1.35), which have huge impacts. With a relatively more significant number of studies included in this synthesis, vocabulary exercise is highly recommendable, including sentence–making and vocabulary worksheets.

Muncie (2002) pointed out that writing in context, with attention to vocabulary use, is a general second language improvement tool. Therefore, as one of the forms ofwriting, sentence making helps participants remember the form of the words and the meaning in specific contexts. Other vocabulary exercises provide participants with opportunities to meet the unfamiliar words repeatedly. In addition, with the instructor's permission and encouragement of dictionary usage, participants are more likely to understand the unknown word immediately, which helps participants read more confidently and smoothly. However, in some primary studies, the dictionary is not allowed because participants are encouraged to guess the meanings according to the contexts instead of reading intensively.

In the pre-post, comprehension questions (d = 4.19) yield a considerable effect, followed by dictionary usage and others, which have huge effects (d = 3.85). Two methods contained in others, MoodleReader and oral rendition, are reported to be conductive to students' vocabulary learning. Robb (2005) argues that MoodleReader is also an extensive reading promoting program that allows students to do randomized vocabulary quizzes on Graded Readers online. As one of the audio–assisted reading methods, oral rendition provides participants with opportunities to learn both written and spoken forms of a word (Chang, 2013). Moreover, the book report also reveals its positive effect, involving reading journals, logs, and book summaries. It thus reflects that unfamiliar words encountered in extensive reading are more likely to be acquired through writing the outline of the plot and reflection of the book.

6. Conclusion

Due to the fact that the overall strength of extensive reading resulted in a very large effect size on group contrasts (d = 1.32) and massive effect size on pre–post contrasts (d = 3.26), we can state categorically that extensive reading has a significant effect on English vocabulary learning. Regarding instruction specifications, it is suggested that teachers adopt Graded Readers as comprehensive reading materials and apply both comprehension questions and vocabulary exercises as teaching methods to promote EFL learners' vocabulary learning. One semester (less than three months) is the most appropriate length of extensive reading instruction for vocabulary learning. Furthermore, extensive reading performs better in helping students memorize vocabulary when the related instruction lasts shorter. Regarding test use and test reliability, the results of studies are more reliable if the same test version is used in the pretest and posttest. The reliability is reported by citing the data from the test manual.

The authors hope this study may provide helpful insights for extensive reading research with the issues in this meta-analysis. Teachers tend to have more new and effective options to promote students' vocabulary acquisition. Moreover, further future studies are encouraged to investigate the specific reasons for the exciting findings revealed in this meta-analysis.

Most of the studies covered in this meta-analysis are university students, followed by children. The length of the studies is primarily short term (less than six months). Therefore, it is suggested that there should be more research with other age groups over one year of instruction. Furthermore, more studies of a large sample sizeregarding the effects of extensive reading on vocabulary should be included in the future meta-analysis. It is more comprehensive and conductive to adopt meta-analysis as the methodology of research synthesis than any independent study due to its revealing of the past research tendency and

future research direction.

References

Alahirsh, H. (2014). *Exploring the effectiveness of extensive reading on incidental vocabulary acquisition by EFL learners: An experimental case study in a Libyan University* (Doctoral dissertation, University of Nottingham).

Alavi, S., & Keyvanshekouh, A. (2012). Using the MoodleReader as an extensive reading tool and its effect on Iranian EFL students' incidental vocabulary learning. *English Language Teaching*, *5*(6), 135-145. https://doi.org/10.5539/elt.v5n6p135

Al-Homoud, F. A., & Alsalloum, M. S. (2012). The effects of extensive reading on the breadth and depth of vocabulary knowledge and reading speed. In *Extensive Reading World Congress Proceedings*, *1*(6), 5-67.

Al-Homoud, F., & Schmitt, N. (2009). Extensive reading in a challenging environment: A comparison of extensive and intensive reading approaches in Saudi Arabia. *Language Teaching Research*, *13*(4), 383-401. https://doi.org/10.1177/1362168809341508

Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese university EFL learners'reading rates. *Language Learning*, *62*(3), 665-703. https://doi.org/10.1111/j.1467-9922.2011.00651.x

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words forprimary grade vocabulary instruction.

Cha, J. E. (2009). The effects of extensive reading on enhancing vocational high school students' L2 vocabulary& reading rates. *English Teaching*, *64*(3), 3-30. https://doi.org/10.15858/engtea.64.3.200909.3

Chang, A. C. (2015). An investigation of different text levels on L2 learners' vocabulary learning rates in an extensive reading program. *Vocabulary Learning and Instruction*, *47*. 47-57.

Chang, Anna C–S. (2013). The effect of audio–assisted extensive reading on vocabulary learning and retention.
Extensive Reading World Congress, 299-310.

Chin-Neng, C. H. E. N., Shu-Chu, C. H. E. N., Shu-Hui, E. C., & Shyh-Chyi, W. E. Y. (2013). The effects of extensive reading via e-books on tertiary level EFL students' reading attitude, reading comprehension and vocabulary. *TOJET: The Turkish Online Journal of Educational Technology*, *12*(2).

Cobb, T. (2007). *Computing the vocabulary demands of L2*

reading. Cohen, J. (1988). *Statistical power for the behavioral*

sciences. Hillside.

Day, R. R. (2004). *Extensive reading activities for teaching language*. Ernst Klett Sprachen.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation ofresearch results* (pp. 10-11). Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, *5*(10), 3-8. https://doi.org/10.3102/0013189X005010003

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice* (p. 328).

Ernst Klett Sprachen.Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003).

Measuring inconsistency in meta-analyses.
BMJ: British Medical Journal, *327*(7414), 557. https://doi.org/10.1136/bmj.327.7414.557

Hong, S. (2014). *Extensive reading in a second language: Literature review and pedagogical implications*
(Doctoral dissertation).

Ioannidis, J. P., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: A large survey. *Canadian Medical Association Journal*, *176*(8), 1091-1096. https://doi.org/10.1503/cmaj.060410

Ismael, H. I., Mohamed, H. A., & Mohammed, M. A. (2017). *Investigating the role of extensive reading in developing student speaking skill* (Doctoral dissertation, Sudan University of Science and Technology).

Jafarpour, M. (2014). The effect of incorporating the extensive reading approach into vocabulary instruction on learners' lexical depth in an EFL context. *Advances in Language and Literary Studies*, *5*(3), 162-169. https://doi.org/10.7575/aiac.alls.v.5n.3p.162

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis.
The modern language journal, *73*(4), 440-464. https://doi.org/10.1111/j.1540-4781.1989.tb05325.x

Kweon, S. O., & Kim, H. R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a foreign language*, *20*(2), 191-215.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33-51. https://doi.org/10.1177/026553229901600103

Lee, S. Y. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal*, *38*(2),150-170. https://doi.org/10.1177/0033688207079730

Light, R., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Harvard

University Press. Lin, L. F. (2010). Senior high school students' reading comprehension of

graded readers. *Journal of Language*
Teaching and Research, *1*(1), 20-28. https://doi.org/10.4304/jltr.1.1.20-28

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

Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles acrossdrafts. *System*, *30*(2), 225-235. https://doi.org/10.1016/S0346-251X(02)00006-4

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. *Handbook of reading research*, *3*, 269-284.

Nakanishi, T. (2015). A meta–analysis of extensive reading research. *Tesol Quarterly*, *49*(1), 6-37. https://doi.org/10.1002/tesq.157

Nation, I. S. P. (2001). *Learning vocabulary in Another Language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

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

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00 4769), Washington, DC: U.S. Government Printing Office.

Palmer, H. E. (1964). *The principles of language study* (p. 111). Oxford University Press.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-languagevocabulary learning. *Journal of Memory and Language*, *30*(3), 331-347. https://doi.org/10.1016/0749-596X(91)90040-Q

Rafatbakhsh, E., & Alavi, S. (2013). The impact of computerized and paper-based "whodunits" as extensive reading materials on Iranian EFL learners' incidental vocabulary learning. *Extensive Reading World Congress 2.* (pp. 284-298).

Rashtchi, M., & Pourmand, S. (2014). Controlled extensive reading and vocabulary knowledge: Let's move towards autonomous learning in EFL contexts. *International Journal of English and Education*, *3*(2), 132-146.

Rezaee, A. A., & Nourzadeh, S. (2011). Does extensive reading improve EFL learners' processing ability?
Theory and practice in language studies, *1*(9), 1167-1175. https://doi.org/10.4304/tpls.1.9.1167-1175

Ricahrds, J. C., & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics* (3rded.).

Robb, T. N. (2005). Outline of MoodleReader features. Retrieved December 10, 2010, fromhttp://moodlereader.org/moodle/mod/resource/view.php?inpopup=true&id=3521

Rodrigo, V., Krashen, S., & Gribbons, B. (2004). The effectiveness of two comprehensible-input approaches toforeign language instruction at the intermediate level. *System*, *32*(1), 53-60. https://doi.org/10.1016/j.system.2003.08.003

Rosszell, H. R. (2007). *Extensive reading and intensive vocabulary study in a Japanese university*. TempleUniversity.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention,assessment and adjustments*. John Wiley & Sons.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2),467-474. https://doi.org/10.22237/jmasm/1257035100

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions ofthe vocabulary levels test. *Language testing*, *18*(1), 55-88. https://doi.org/10.1177/026553220101800103

Soltani, R. (2011). Extensive reading: A stimulant to improve vocabulary knowledge. *Studies in literature andlanguage*, *2*(3), 161-167. h

Stahl, S. A. (2005). Four problems with teaching word meanings. *Teaching and learning vocabulary: Bringingresearch to practice*, 95-114.

Sterne, J. A., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *Stata Journal*, *4*, 127-141.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on

choice of axis.
Journal of clinical epidemiology, *54*(10), 1046-1055. https://doi.org/10.1016/S0895-4356(01)00377-8

Suk, N. (2017). The effects of extensive reading on reading comprehension, reading rate, and vocabulary acquisition. *Reading Research Quarterly, 52*(1), 73-89. https://doi.org/10.1002/rrq.152

Tabata-Sandom, M. (2017). L2 Japanese learners' responses to translation, speed reading, and 'pleasure reading'as a form of extensive reading. *Reading in a Foreign Language, 29*(1), 113.

Tan, D. A. L. (2016). *Investigating guided extensive reading and vocabulary knowledge performance among remedial ESL learners in a public university in Malaysia* (Doctoral dissertation, University Sains Malaysia).Teng, F. (2015). Extensive reading plus explicit vocabulary exercises: Is it better than extensive reading-only?
Malaysian Journal of ELT Research, 11(2), 82-101.

Teng, F. (2015). The effectiveness of extensive reading on EFL learners' vocabulary learning: Incidental versus intentional learning. *BELT-Brazilian English Language Teaching Journal*, *6*(1), 66-80. https://doi.org/10.15448/2178-3640.2015.1.20033

Tiryaki, Z. M., & Tütüniş, B. (2012). The role of extensive reading on vocabulary. *ELT Research Journal*, *1*(4), 209-215.

Vaezi, R., & Nilforooshan, S. (2013). The effect of using Moodle Reader as an extensive reading tool on learners' awareness of affective dimensions of deep vocabulary knowledge. *Mediterranean Journal of Social Sciences, 353*.

Waring, R. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader?
Reading in a Foreign language, *15*(2), 130.

Webb, S., & Chang, A. C. S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, *37*(4), 1-25. https://doi.org/10.1017/S0272263114000606

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth.
Canadian Modern Language Review, *53*(1), 13-40.

Wijaya, H. P. S. (2013). *Extensive reading program: Its impacts on student's vocabulary size and readingcomprehension* (Doctoral dissertation, Petra Christian University).

Yamamoto, Y. (2011). Bridging the gap between receptive and productive vocabulary size through extensivereading. *The Reading Matrix, 11*(3), 226-242.

Yamashita, J., & Kan, K. N. (2010). Examining effects of L2 extensive reading in the cognitive and affective domains. *Applied Linguistics, Global and Local*, 375-386.

Zhang, T. S. (2015). *Advanced meta-analysis methods: Based on Stata* (p. 36). Fudan University Press.