The Relationship between the Amount of Extensive Reading and the Writing Performance

Nobuko Sakurai

Kyoto Sangyo University

ABSTRACT

This paper explored the effects of the amount of extensive reading (ER) on writing ability. Participants were 157 first and second-year non-English majors at a private university in Japan who took a writing test in class. Some of them were reading extensively, while others had no experience in ER. The outcomes of Pearson's correlation indicated that ER correlated with total writing scores positively. A one-way analysis of variance with a planned comparison was performed with the total number of words read as the independent variable and total scores and scores for four areas of writing, task achievement, coherence and cohesion, lexical resource and language use, as dependent variables. The results showed that the amount of reading made a statistically significant difference in the mean scores of vocabulary and grammar. Moreover, findings suggested that participants who had accomplished reading more than 108,000 words statistically significantly wrote better lexically and grammatically.

INTRODUCTION

In this age of advanced technology, writing seems to be a prominent means of communication. Convenient devices such as smartphones and tablets have enabled people to send emails and post messages onto Social Networking Service sites anytime anywhere. Although whether or not text messages and tweets are considered as writing is debatable, it is de facto that they are a way to communicate in place of speaking. Writing, more than or as much as speaking, has become part of most people's daily lives.

LITERATURE REVIEW

Writing in the Japanese English education

Writing is not only spelling words and putting words together. In addition to vocabulary, grammar, mechanics and organization, it involves content, purpose and audience (Raimes, 1983). Therefore, writing cannot be acquired naturally. Even though humans start speaking their native language without formal instructions, they need to learn to write at school (Raimes, 1983). Students have to be taught how to organize sentences, paragraphs and ideas coherently (Harmer, 1991).

However, writing has not been the center of English education in Japan. Most Japanese students officially begin learning English at junior high school and continue studying it at senior high school. These six years of English learning are greatly influenced by the Course of Study, guidelines issued by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). It specifies how much of what should be taught in what stage of education from kindergarten to senior high school. The one for junior high school English that was revised in 2008 and is currently followed states that English should be taught integratively for communication (MEXT, 2008). With regard to writing, it notes that students should be able to write their ideas and feelings about events and experiences in familiar situations by connecting words and sentences correctly. How well this objective is being achieved can be examined in textbooks since MEXT screens textbooks and approves only those that are written in accordance with the guidelines. In other words, the national education policy is reflected in textbooks. Teachers are well aware of this system, so they tend to believe following textbooks closely is equal to following the guidelines (Wada, 1997, cited by Sakurai, 2007). For example, the lesson plans prepared by the publisher for Book 3 of *New Horizon*, one of the popular junior high school English textbooks, suggest that only three lessons throughout the year focus on writing (Tokyo Shoseki, 2015). Although there are other activities that require writing, they are regarded as preparation for speaking. According to Raimes (1983), writing is different from speech put down on paper. This implicates that writing is not instructed sufficiently at junior high school.

Writing lessons also appear to be lacking in the next three years of formal education after junior high school. The Course of Study for English at senior high school outlines one required course and six elective courses, and the description of these courses implies that little time is spent on writing (MEXT, 2009). In fact, the outcomes of a four-skills test and nationwide questionnaires conducted officially by the education ministry in 2014 proved this (MEXT, 2015). Participants of the study were 70,000 randomly chosen third-year senior high school students and their English teachers. Students sat for reading, listening and writing tests,

and 17,000 among them also took a speaking test. The results confirmed that their writing skills as well as speaking needed more improvement. Furthermore, the survey results disclosed the fact that a lot of teachers were not conducting enough speaking and writing activities.

Extensive reading (ER)

ER is defined as reading a lot of easy, enjoyable books at a good speed. A great deal of research has acknowledged that ER has a positive impact on various aspects of language development. Krashen (2004) and Day and Bamford (1998) summarized numerous studies that proved the efficacy of ER on vocabulary, grammar, spelling, listening, test scores, motivation to learn English, not to mention reading ability.

Learners in the English as a foreign language environment are expected to be exposed to a great deal of input by engaging in ER. It has been gaining popularity in Japan. Large bookstores often have an ER section, and major publishers specializing in English language teaching print an exclusive catalog of readers separate to their catalog for textbooks.

ER and writing

Some researchers have focused on how attributable ER is to writing performance and writing-related skills in English. To quote a few to start, Janopoulos (1986) investigated the number of hours that 79 graduate students of mixed nationalities in the US spent on pleasure reading weekly when administering a writing test that was later holistically graded. He found a correlation between reading in the second language (L2) and writing in L2. Lee (2005) examined factors that were considered to enhance or impede writing, and identified ER as the only statistically significant predictor for writing performance among writing apprehension, writer's block, free reading and free writing.

The most common methodology to study effects of ER on writing appears to be analytically marking an essay after the treatment of ER and comparing the score to that of a pre-reading writing test. It was reported that 25 Pakistani secondary school students made gains in fluency, vocabulary and accuracy of expression after reading extensively for 23 weeks (Hafiz & Tudor, 1990). Lai (1993) stated that 52 secondary school students in Hong Kong who read 14.2 books on average statistically significantly increased the number of words on the post-writing test. He also noted that participants improved spelling and verb agreement. Moreover, research conducted in Hong Kong by Tsang (1996) revealed that the ER group improved content, language use and the overall quality of writing over the period of 24 weeks compared to the control group and the group that received writing-oriented lessons. Findings by Lee and

Hsu (2009) showed that 43 Taiwanese vocational college students who read extensively in class for one year outperformed the students in the control group on the post-writing assignment. Improvement was statistically significant in all areas: fluency, content, organization, vocabulary, language use and mechanics. Mermelstein (2015) carried out similar research with 107 university students in Taiwan. Participants who did ER in and outside the classroom achieved significant gains in content, vocabulary, language use, spelling and mechanics, and fluency on the essay test administered after the treatment of ER for 29 weeks. A recent study by Park (2016) yielded similar results. English as a second language learners preparing to be undergraduate and graduate students in the United States engaged in Sustained Silent Reading (SSR) for 15 minutes over 32 lessons in an intermediate writing course. They were instructed to continue reading and finish a 10-minute long writing activity as homework every week, and they discussed their books for 5 minutes in the following lessons. These students in the experimental group did better in the areas of content, organization, vocabulary and language use on a post essay writing test than students in the control group.

Unfortunately, all these six studies mentioned in the previous paragraph and compiled in Table 1 do not discuss the ER books utilized in their research to a great extent. It is not feasible to judge what reading materials were employed in the study by Tsang (1996). Although Hafiz and Tudor (1990), Lai (1993), Lee and Hsu (2009), Mermelstein (2015) and Park (2016) mentioned the series names of books, not all of them specified the levels of the series. There are various types of books that can be provided in ER programs. Graded readers (GRs), leveled readers (LRs) and picture books are some examples. They are written for a different target group, so language used, especially vocabulary, varies. Even in the same type, each series often has unique characteristics. Similarly, levels of books make a difference. Generally speaking, higher-level books tend to include classic literature, and their plots are more complicated. As reading materials are a source of input for readers, the quality of ER books supplied affect students' learning.

Most crucially, the exact reading amount is not clear or unknown in all of these studies. Lai (1993) measured the reading amount in the number of books, while Lee and Hsu (2009) in the number of pages. However, how many pages in a book and how many words on a page there are differ greatly depending on the series and levels of books. Take books in Oxford Bookworm series as an example. "A Little Princess" (level 1) contains 41 pages, whereas there are 104 pages in "Cry Freedom" (level 6). One full page without illustrations, headings and break lines contained 260 words (p. 24) in "A Little Princes" and 297 words (p. 103) in "Cry Freedom," respectively. Also, the mean running words range from 5,641 to 29,408 between level 1 and level 6 books (Furukawa, Kanda, Mayuzumi, Nishizawa, Hatanaka, Satoh, & Miyashita, 2013). It is also challenging to estimate the amount of reading from the number

of hours spent on ER. The reading rate of L2 learners during ER varies from research to research as summarized by Sakurai (2015). Psychological and physical factors such as state of mind and tiredness could cause reading speed to fluctuate (Sakurai, 2015). Differences that could be triggered by the number of words read together with types, series and levels of ER books should not be ignored.

<Insert Table 1 here>

RESEARCH QUESTIONS

Based on the results of studies, ER is speculated to contribute to the improvement of writing skills. This means that the amount of ER could indicate learners' writing ability. Krashen (2004) said that we learn to write by reading. If this is true, the amount and quality of input that reading materials provide should matter. It seems unmanageable to perfectly reflect the quality of input on research since one of the beneficial features of ER is that readers get to choose what they want to read. Individual preferences are not feasible to measure. On the other hand, the exact amount of reading recorded in the number of words rather than the number of pages, books or hours spent can be taken into account. Consequently, the research questions under investigation are:

1. Can the number of words read extensively indicate writing ability?
2. If so, what sub-skills of writing does it affect?

METHOD

Participants

Participants of the present study were 157 first and second-year students at a private university in Japan. To be more specific, they were 52 Asian language, 89 European language and 16 science majors who were enrolled in an English program. All the courses offered in the program were elective. However, these students needed a certain number of credits from the program to graduate or to complete a special certificate.

Writing test administered in the English program

The program is special in the way that it provides various types of courses (i.e. skill-based, content-based, seminar-style, lecture-style, one-credit, two-credit, intensive during

summer and spring holidays, and summer camp). Even though it is organized and managed by the English Department of the Faculty of Foreign Studies, it is open to any undergraduate student of any faculty on campus.

There are courses of four different levels in the program: elementary, intermediate, intermediate to advanced, and advanced. Previously, TOEIC, TOEIC Bridge and TOEFL scores were mainly used to place students in the right level. However, a lot of instructors in the program pointed out the gap in the speaking and writing abilities between the participants and English and International Relations (IR) majors, who were intensively studying English every weekday as their first foreign language. Participants of the current study, on the other hand, were taking only two English classes a week, a TOEIC and a communication-based class, in the general education program. Accordingly, it is reasonable to say that the productive skills of English and IR majors were higher than that of the participants.'

Requests and suggestions by the instructors and students who noticed the discrepancy between the test scores and the actual abilities to speak and write led to the implementation of the writing and speaking tests original to the program. The main objective was to label students with an intermediate level score in reading and listening but without sufficient writing and speaking skills as elementary. It was decided that students with an intermediate level score in one of the three aforementioned official English proficiency tests should proceed to the writing test, and that those who reached the intermediate level in writing were allowed to take the speaking test.

Instruments

The writing test given in the program was letter writing. In the English program, there were four writing courses: SNS English (sentence writing), Email English (paragraph writing), Essay Writing and Academic Writing. Writing ability of the majority of participants was assumed to be elementary. The Essay Writing course was for intermediate students, while Email English that dealt with letter writing was for elementary. This led to the decision that the test should not be an essay but a letter. Another deciding factor was that major English proficiency tests such as International English Language Testing System (IELTS) include a letter writing task.

The task was to write a letter of at least 150 words long to a non-Japanese friend abroad about recent news in Japan and explain why Japanese people were interested in it (see Appendix A). The test was conducted in November 2015, for 30 minutes in 32 intermediate, intermediate to advanced, and advanced level courses.

Evaluation criteria were prepared beforehand with reference to the IELTS scoring

system and the scales in the Common European Framework of Reference for Languages. They included task achievement, coherence and cohesion, lexical resource and language use (see Appendix B). The scores for each sub-skill ranged from 0 to 4 with 16 as a full mark. After some training, five full-time instructors graded the writing tests as a first rater. The author of the current study marked all the writing as the second rater.

Most of the intermediate to advanced and advanced level courses in the program had an ER component as 20% of the final grade. Students were required to read a minimum of 36,000 words to pass the course. As a rule, they individually selected titles from among more than 12,000 ER books in the library, read them outside the classroom, took quizzes on MReader and accumulated the number of words that the books contained. The majority of students started ER in this manner, while the others had taken a 15-week long course called Introduction to ER before reading independently. In this course, students experienced SSR for entire lessons keeping a reading record sheet where they wrote a brief summary of books. The number of words was accumulated when the teacher admitted that the summary was satisfactory. Also, some participants were enrolled in a course which demanded them to accumulate a minimum of 75,000 words by doing ER with audio in class and with MReader outside the classroom. The reading amount accounted for 50% of the final grades in these two ER-oriented courses.

Procedure

Each score for the four areas of writing by the first and second raters were input into Excel and the total scores were computed. Then, a one-way analysis of variance (ANOVA) was conducted between the raters to see if all six instructors evaluated the writing in line with the scoring rubric. The total number of words students read extensively by the end of the semester together with the record of quizzes they passed were downloaded from MReader. Then, the amount of reading done by participants who had taken the Introduction to ER course was estimated from their final grades as shown in Table 2, and it was added to the MReader record. Following descriptive statistics, Pearson's correlation coefficients were carried out in order to explore how the reading amount was related to the writing scores. A one-tailed test was adopted as the relationship between the independent variable and dependent variables was hypothesized to be one-directional (Field, 2009). After that, participants were divided into four groups based on the total number of words they accumulated by reading extensively. The first group consisted of students who read less than 36,000 words and those who had no ER experience. Students with 36,000 to 71,999 words constituted the second group, and those with more than 72,000 but fewer than 107,999 the third group. The last group comprised of students who accumulated more than 108,000. In other words, participants who accomplished the minimum

ER requirement of one intermediate to advanced or advanced level class in the English program were in the second group and two in the third, and three in the forth. ANOVA with a planned comparison was run between these four groups with the reading amount as the independent variable and the scores for the four subs-skills as well as the total scores as dependent variables. All the analyses were performed utilizing SPSS (version 21.0).

Table 2. Estimated number of words read in Introduction to ER indicated by the final grades

Final grade (points)	Estimated number of words read (words)
90 – 100	25,000
80 - 89	20,000
70 - 79	15,000
60 – 69	10,000

RESULTS

Inter-rater reliability

In order to validate the writing test scores for analyses, ANOVA was calculated among the six raters with the scores for the four areas of writing and the total scores as dependent variables. The results were significant at the $p < .001$ level, $F (5, 610) = 19.048$ for task achievement, $F (5, 610) = 6.982$ for cohereance and cohesion, $F (5, 610) = 22.939$ for lexical resource, $F (5, 610) = 17.501$ for language use, $F (5, 610) = 12.540$ for the total. This unfortunately proved that the scorers were evaluating the tests somewhat differently, and that the data were subsequently as unreliable as they were. Thus, it was determined that the scores by the first raters were disregarded and only the data from the second scorer were going to be utilized for further analyses. This appeared to be reasonable according to the report by Zhang, Xiao and Luo (2015). They stated that the reliability of one rater was close to 0.8 for informal, everyday evaluation.

Descriptive statistics

Table 3 demonstrates the descriptive statistics of the number of words participants read extensively, total scores and scores for each subscale of the writing tests. No one received the lowest score of 0 or highest score of 4 in any of the four areas of writing. The scores for task achievement, coherence and cohesion and language use ranged from 1.0 to 3.5. The mean

scores of these three sub-skills (1.885 for task achievement, 1.825 for coherence and cohesion and 2.392 for language use), however, suggest that students did best on language use among the three. Vocabulary seems to have been where test takers were able to achieve the most as the minimum score (1.5) and the mean (2.513) were the highest among the four. The average total score was 8.615 with 12.5 points as the maximum and 5.5 as the minimum. The mean number of words participants read was 66,242.47. Surprisingly, one Asian language major accumulated 350,480 words in two years.

Table 3. Descriptive statistics of the reading amount, scores for the four sub-skills and the total scores on the writing tests

	N	Min	Max	Mean	SD
Number of words read	157	0	350,480	66,242.47	62,405.849
Task achievement	157	1.0	3.5	1.885	.6526
Coherence & cohesion	157	1.0	3.5	1.825	.6647
Lexical resource	157	1.5	3.5	2.513	.4769
Language use	157	1.0	3.5	2.392	.5290
Total	157	5.5	12.5	8.615	1.6860

Pearson's correlation (one-tailed)

The hypothesis that more ER resulted in better scores on the writing tests was considered reasonable. Therefore, a one-tailed correlation test was applied. As presented in Table 4, the amount of reading statistically significantly correlated with the total scores and scores for the two sub-skills except for task achievement and coherence and cohesion. The ER amount had the strongest correlation with lexical resource at the $p < .001$ level, and with language use and the total at the $p < .01$ level.

Table 4. Results of Pearson's correlation

	TA	C & C	LR	LU	Total
ER	.010	.122	.294***	.246**	.212**

***= $p < .001$, ** = $p < .01$, * = $p < .05$

Note: TA = task achievement, C&C = coherence and cohesion, LR = lexical resource, LU = language use

ANOVA

Participants were divided into 4 groups based on the amount of ER. Because most of them were in courses with the requirement of reading at least 36,000 words, groups were decided depending on how many classes worth students read so far. Details of the groups are shown in Table 5.

A one-way ANOVA with a planned comparison was run between the four groups with the five scores from the writing tests as dependent valuables. Table 6 demonstrates the results of ANOVA and Table 7 summarizes the outcomes of the planned comparison. Findings show that there was a statistically significant difference in the mean score of lexical resource, F (3, 153) =5.445, $p < .01$, and of language use, F (3, 153) =4.057, $p < .01$. The outcomes of the planned comparison revealed that the group that read more than 108,000 words statistically significantly outperformed the other groups in the area of lexical resource at the $p < .001$ level (t = -3.764). Moreover, the mean score of language use was statistically significantly different at the $p < .01$ level between participants who read more than 108,000 words and who read fewer (t = -3.162).

Table 5. Descriptive statistics of the four groups

Group	N	Reading amount (words)	Mean	SD	# of classes
1	51	0 – 35,999	7,640.78	10,237.988	0
2	42	36,000 – 71,999	48,680.14	10,525.817	1
3	31	72,000 – 107,999	84,350.74	10,635.139	2
4	33	108,000 –	162,149.97	53,782.0273	3

Table 6. Results of ANOVA

Source of variance	Lexical resource			
	SS	df	MS	F
Between groups	3.422	3	1.141	5.445**
Within groups	32.053	153	.209	
Total	35.475	156		

Source of variance	Language use			
	SS	df	MS	F
Between groups	3.217	3	1.072	4.057**
Within groups	40.442	153	.264	
Total	43.659	156		

***= $p < .001$, **= $p < .01$, * = $p < .05$,

Table 7. Results of the planned comparison

Comparison	Groups	Lexical resource		Language use	
		df	*t*	df	*t*
1	1 - 3 vs. 4	153	-3.764***	153	-3.162**
2	1 - 2 vs. 3	153	-1.061	153	-.553
3	1 vs. 2	153	-.279	153	-1.066

***= $p < .001$, **= $p < .01$, * = $p < .05$,

ER books read by participants

The record of quizzes that students passed to accumulate words was downloaded from MReader. It was expected that it would suggest the quality of input they were internalizing. The names and levels of the series which quizzes were passed more than 10 times by participants overall are listed in Appendix C. It also presents the number of times quizzes were taken by participants in each group.

DISCUSSION

Research question #1

The first research question asked if the number of words read extensively could suggest writing ability. It is appropriate to conclude that the answer to this question is not affirmative. While the reading amount correlated with the total scores at the $p < .01$ level, the results of ANOVA were not significant. One possible interpretation is that ER influences some sub-skills of writing, but the effect is not remarkable enough to affect the total.

Research question #2

The second research question inquired what areas of writing ER contributed to, and they are vocabulary and grammar. From the outcomes of the planned comparison, it is interpreted that ER had a favorable impact on lexical and grammatical development. Reading more than 108,000 words resulted in the significantly better mean score of lexical resource at the $p < .001$ level, and of language use at the $p < .01$ level. It is speculated from findings of this study that vocabulary, among various aspects of language that ER positively attributes to, may be an area where great improvement can be expected.

It is salient that what was observed in the current study seems to be concerned with memorized lexical items in the past being used in production. A great deal of research on vocabulary learning through reading has been published. Nonetheless, often-cited studies were designed to measure the receptive knowledge of vocabulary irrespective of the retention rate, and they failed to investigate the reproductive rate (Waring & Takaki, 2003). In fact, all of the research complied by Waring and Takaki (2003) focused on new vocabulary tentatively learned through reading. For example, in the study by Day, Omura and Hiramatsu (1991), Japanese senior high school and university students took a multiple-choice vocabulary test where they needed to identify the correct definition or antonym of new words that appeared in a slightly modified version of a mystery that they read in class. Japanese junior and senior high school students are often directed to make a vocabulary list or to buy a vocabulary book that has English words on the left and Japanese translation on the right. They then memorize as many words as possible with the list or book for quizzes, term and entrance exams. Meeting the words that they had studied in their previous education during ER, participants might have recognized the usage of them, and it might have resulted in a better writing performance. Passive vocabulary could have been developed into active vocabulary due to ER.

Participants chose to read more GRs than LRs as shown in Appendix C, and GRs may have played a crucial role in the transition of lexical knowledge to production. As Waring and Nation (2004) stated, ideal vocabulary coverage is supplied by GRs. Waring and Nation (2004) cited the conclusion drawn by Nation & Wang (1999) that learners should be encouraged to read at least one GR per week in order for their learning to be reinforced. This is because one GR every week causes readers to encounter the same word more than 20 times, and it can ensure that the word is transferred from the short-term to the long-term memory. It could be said that 108,000 words is equivalent to one weekly GR, and reading this much amount is necessary for the memory process to be completed.

It was also confirmed that students with more than a 108,000 word count in ER gained more points in language use. It is inferred that a variety of grammatical structures were accurately used on the writing tests. This can be analyzed with the same theory that accounted for the gains in lexical resource. ER facilitated grammar knowledge acquired during six years before university to be digested and activated.

It may be worth comparing books participants who accumulated more than 108,000 words read to those selected by the others. It is noticeable that the books students in Group 4 read ranged from short, easy books to longer, more challenging books, while those in the other three groups were not reading as many starter-level books. To be more precise, participants with fewer than 108,000 words neglected to read Building Blocks Library (BBL) levels 5 to 7, and students in Groups 2 and 3 took quizzes for BBL level 8. The same pattern is seen with the

Foundations Reading Library series. Cambridge English Readers Starter level, Compass Young Learner's Classics levels 1 to 4, Oxford Classic Tales level 1, Oxford Reading Tree levels 5 to 7, and Penguin Readers level 0 were read more by learners in Group 4. Similarly, these students took more quizzes for Oxford Bookworms level 0 compared to students in Group 1. It is natural that students who accumulated more than 108,000 words read more books than those with fewer words. Still, it is assumed that starting with books of fundamental levels and gradually climbing up on a ladder could be essential for the transition of lexical and grammatical knowledge into practice.

All of the six studies described in Table 1 yielded similar results. Descriptive essay writing was assigned as pre and post- tests in these studies, while the task was letter writing in this study. This might mean that the style of writing does not make a difference.

The obtained F-values of task achievement as well as of coherence and cohesion were not found significant. The task of the writing tests was to write a letter to inform a foreign friend of news in Japan and discuss why it was interesting to Japanese people. It required some knowledge of the current world that fiction-oriented ER books did not draw on. Two explanations can be offered for the insignificant outcomes in coherence and cohesion. First, as previously mentioned, Japanese students hardly learn how to write a paragraph in English at junior and senior high schools. Writing is merely a step before a speaking activity. Therefore, it is assumed that they basically translate Japanese into English sentence by sentence and that the teacher is not teaching how to organize sentences and ideas. Second, a lot of participants read more GRs than LRs. GRs are written with a limited kind and number of vocabulary, and grammar used in them is restricted. Consequently, some stories such as a shortened version of a long, famous story and a story based on a movie are rather abrupt and unnatural. Hafiz and Tudor (1990) appeared to have admitted the importance of the quality of input. They attributed the significant gains in the accuracy of grammar to the book selection that was easy and simple enough for their students. Lack of experience in writing organized paragraphs and the probable influence of GRs could have inhibited students from attaining a sense of the natural flow of writing.

LIMITATIONS

There are limitations to the current study. First, the research design was different from previous studies in which essays were evaluated as pre and post-tests. In this study, instead, the letter writing tests were administered in order to distinguish students with the intermediate level of writing ability from those without. In other words, the tests were not for research purposes. The results did not demonstrate improvement of writing skills after the treatment of

ER. Rather, they merely indicated writing skills of learners with a certain amount of ER. Also, the writing tests should have included more than one task in order to stabilize data.

Secondly, scores by two raters should have been utilized. However carefully evaluation criteria were followed, data could have been skewed by one rater. The six raters needed more training. Analytic scoring was employed in the present study rather than holistic scoring. Two full-timers completed a scoring rubric and one of them led a training session for the rest of the scorers where all of them marked the same samples together and discussed how they decided on the ratings. This certainly helped them stand on the same ground, but it must have been insufficient. According to Bauer (1981) cited in Zhang et al. (2015), analytic scoring requires twice as much training compared to holistic scoring. Another point is that perception of the evaluation criteria seemed to have slightly differed among the raters. To be more specific, they had to judge if test takers achieved the task of writing about interesting news in Japan and explaining why it intrigued Japanese people. In the reflection session, some scorers commented that they accepted beautiful foliage in season as news since it was being reported on TV news programs at that time. On the other hand, the other teachers considered it inappropriate, as foliage is a natural yearly phenomenon. It is possible that the scores for task achievement could not have been legitimate.

Furthermore, the scoring method might have also influenced the results. Tsang (1996), Lee and Hsu (2009), and Mermelstein (2015) followed the criteria invented by Jacobs, Zinkgraf, Wormuth, Hartfield and Hughey in 1981 or adopted an arranged version of them. However, Hafiz and Tudor (1990), Lai (1993) and the author of this study evaluated writing according to their own set standard. Scoring methods have not been standardized across the board, and each examiner or researcher can decide on evaluation criteria that serve the purpose of tests best. How writing is graded produces a direct consequence in the score.

Finally, participants could have been grouped differently. For the purpose of improvement in the English program, groups were determined based on the reading amount necessary to pass one to three courses. This grouping may not have been the best for a general use of the results.

CONCLUSION

It is not plausible to conclude that the amount of ER could indicate writing ability of learners   (research question #1). Nevertheless, it was found that students who read more than 108,000 words outperformed those who read fewer in the areas of vocabulary and grammar (research question #2). As Brown, Waring and Donkaewbua (2008) stated, ER must have enabled participants to internalize lexical and grammatical knowledge that was previously

learned.

Writing ability takes time to develop and improve. In fact, even in the famous "book flood" study carried out in Fiji by Elley (1991) that showed significant gains in many aspects of English proficiency, students needed to keep reading for two years before the researcher could confirm gains in writing performance. Nonetheless, it is encouraging that the results of the present study implied that learners who had read more than 108,000 words, which is easily attainable, could write with an adequate use of enriched vocabulary and complex grammar.

Findings of the current study may not be suitable to be referred to in other contexts because of the limitations mentioned above. It is hoped that more research on the efficacy of ER on writing ability using the number of words read will follow.

Nobuko Sakurai *is associate professor in the Faculty of Foreign Studies at Kyoto Sangyo University in Kyoto, Japan. Her research interests include effects of extensive reading on language and cognitive development and on the English educational system in Japan, and of translation in extensive reading programs.*

Email: nsakurai@cc.kyoto-su.ac.jp

REFERENCES

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136-163.

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

Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language, 7*(2), 541-551.

Elley, W. B. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning, 41*(3), 375-411. DOI: 10.1111/j.1467-1770.1991.tb00611.x

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications Inc.

Furukawa, A., Kanda, M., Mayuzumi, M., Nishizawa, H., Hatanaka, K., Satoh, M., & Miyashita, I. (2013). *Eigo tadoku kanzen gaido bukku* (4th ed.) [The complete book guide for extensive reading]. Tokyo: Cosmopier.

Hafiz, F. M., & Tudor, I. (1990). Graded readers as an input medium in L2 learning. *System, 18*(1), 31-42.

Harmer, J. (1991). *The practice of English language teaching.* Essex: Longman.

Janopoulos, M. (1986). The relationship of pleasure reading and second language writing proficiency. *TESOL Quarterly, 20*, 763-768.

Krashen, S. D. (2004). *The power of reading: Insights from the research* (2nd. ed.). CT: Libraries Limited.

Lai, F. (1993). The effect of a summer reading course on reading and writing skills. *System, 21*(1), 87-100.

Lee, S. (2005). Facilitating and inhibiting factors in English as a foreign language writing performance: A model testing with structural equation modeling. *Language Learning, 55*(2), 335-374.

Lee, S., & Hsu, Y. (2009). Determining the crucial characteristics of extensive reading programs: The impact of extensive reading on EFL writing. *The International Journal of Foreign Language Teaching, 5*(1), 12-20.

Mermelstein, A. D. (2015). Improving EFL learners' writing through enhanced extensive reading. *Reading in a Foreign Language, 27*(2), 182-198.

MEXT. (2008). Overview of the course of study for junior high school English. Retrieved March 1, 2017 from http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/ afieldfile/2011/01/05/1234912_010_1.pdf

MEXT. (2009). Overview of the course of study for senior high school English. Retrieved March 1, 2017 from http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/ afieldfile/2010/01/29/1282000_9.pdf

MEXT. (2015). *Seito no eigoryoku kojo suishin puran* [Plans to improve students' English ability]. Retrieved March 1, 2017 from http://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/ afieldfile/2015/07/21/1358906_01_1.pdf

Park, J. (2016). Integrating reading and writing through extensive reading. *ELT Journal, 70*(3), 287-295. DOI: 10.1093/elt/ccv049

Raimes, A. (1983). *Techniques in teaching writing.* Oxford: Oxford University Press.

Sakurai, N. (2007). Changes in high school English education: A comparison of popular English textbooks. *CTA Humanistica Et Scientifica Universitatis Sangio Kyotiensis Humanities Series, 36*, 131-144.

Sakurai, N. (2015). The influence of translation on reading amount, proficiency, and speed in extensive reading. *Reading in a Foreign Language, 27*(1), 96-112.

Tokyo Shoseki, Co., Ltd. (2015). Shido keikaku sakusei shiryo [A guide to lesson planning]. Retrieved March 5, 2017 from https://ten.tokyo-shoseki.co.jp/text/chu/keikaku/eigo/ files/keikaku03.pdf

Tsang, W. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics, 17*(2), 210-233.

Waring, R., & Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English-Speaking World, 4*. Retrieved August 23, 2016, from http://www.robwaring.org/papers/various/waring_120304.pdf.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163.

Zhang, B., Xiao, Y., & Luo, J. (2015). Rater reliability and score discrepancy under holistic and analytic scoring of second language writing. *Language Testing in Asia, 5*(5). Retrieved August 15, 2016, fromhttp://link.springer.com/article/10.1186/s40468-015-0014-4. DOI 10.1186/s40468-015-0014-4

APPENDIX A

Tokubetsu Eigo Fall 2015 English Writing Test

特別英語プログラムレベル分けライティングテスト

All Intermediate and Advanced classes are doing the same test, so you may do it more than once. This is good as it means you will have the chance to practice. We will choose your best answer to judge your level. The result of the test will be used only for the placement and research purposes. （すべての中級・上級・中〜上級クラスで同じテストを行います。２度以上受験することになった場合、一番良い点数を採用します。結果はレベル分けとプログラム開発の研究目的以外の用途には使用しません。）

_____

<u>Evaluation criteria</u>

1. How well you answer the question (task achievement)

2. How well your writing holds together (coherence and cohesion)

3. How good your vocabulary and spelling are (lexical resource)

4. How good your grammar is (grammatical range and accuracy)

<u>Important tips</u>

Spend all the time on the task - and don't rush it. Thirty minutes is not a long time to write a good answer.

Make sure you answer all the parts of the question.

Write in clear paragraphs.

Should your letter be casual, semi-formal or formal? Think about who you are writing to and make sure you use the appropriate level of formality.

Use a good range of vocabulary and go back and check your spelling when you are done.

You will be marked for task achievement, coherence and cohesion, lexical resource and grammatical range and accuracy.

Often you will need to use your imagination to write your letter. Don't just think, "I have no idea." or "I can't." Be creative!

_____

You have **30 minutes** to read the question, plan and write your answer. You should write at least 150 words. You do NOT need to write any address. Begin your letter as follows:

Dear _____

> You have a friend living in another country. They want to know about the news in your country. Write a letter to your friend. In the letter,
>
> - greet your friend and say why you are writing
> - briefly describe a news story in your country
> - explain why people are interested in the news story.

(This question is based on a question from Insearch English *Prepare for IELTS*)

APPENDIX B

	4	3	2	1
Task Achievement	The purpose is generally clear (although there may be some inconsistencies in tone), but the key points of the task are all clearly answered.	Generally addresses the task, although the format may be inappropriate in places. The overall purpose of the letter is clear, although it may be unclear in certain parts. There may be a tendency to focus on details.	Attempts to address the task, but does not cover all key points. May fail to clearly explain the purpose of the letter or the tone may be inappropriate. Parts may be unclear, irrelevant, inaccurate or repetitive.	Seems to have misunderstood the task and so fails to address it. Presents limited ideas which may be largely irrelevant or repetitive.
Coherence & Cohesion	Arranges information and ideas coherently and there is a clear overall progression. Uses cohesive devices effectively, but there may be some mechanical errors.	Presents information with some organization, but there may be some lack of overall progression. Makes inadequate, inaccurate or over-use of cohesive devices. May be repetitive.	Presents information and ideas but they are not arranged coherently and there is no clear progression to the response. Uses some basic cohesive devices but there may be inaccurate or repetitive.	Does not organize ideas logically. May use a very limited range of cohesive devices, but they may not indicate a logical relationship between ideas.
	Uses a good range of vocabulary for the task. Attempts to use less	Uses a relatively limited range of vocabulary, but it is	Uses a basic range of vocabulary which may be used repetitively or	Vocabulary is very limited and spelling and word formation

	common vocabulary mostly accurately and makes some errors in spelling/word formation, but they do not impede communication.	adequate for the task. There are some errors in spelling and word formation that may cause some problems for the reader.	may be inappropriate for the task. There may be noticeable errors in spelling and word formation which cause difficulty for the reader.	errors make comprehension difficult.
Lexical Resource				
Grammatical Range & Accuracy	Uses a mix of complex and simple sentence forms. There are some errors in grammar and punctuation, but they rarely reduce communication.	Attempts complex sentences, but they are less accurate than simple sentences. There are between 5 and 10 grammatical errors.	Uses only a limited range of structures. Punctuation often faulty. More than 10 grammatical errors.	Seems to only be able to use memorized phrases. Frequent grammatical and punctuation errors.

Note. The rubric was compiled in reference to these websites: http://www.smart-words.org/words-for/feelings/happy-glad.html,

http://home.ku.edu.tr/~doregan/Writing/Cohesion.html, http://library.bcu.ac.uk/learner/writingguides/1.33.htm,

https://owl.english.purdue.edu/owl/resource/608/01/, http://www.dcielts.com/ielts-writing/lexical-resource/

http://www2.ivcc.edu/rambo/eng1001/sentences.htm, https://owl.english.purdue.edu/owl/resource/573/02/

APPENDIX C

Quizzes passed more than 10 times by participants overall and the number of times quizzes were passed by participants in each group

Type of books	Series and levels	The number of times quizzes passed				
		Group 1	Group 2	Group 3	Group 4	All
GRs	Building Blocks Library 5	6	1		7	14
	Building Blocks Library 6	6		5	12	23
	Building Blocks Library 7	8	5	8	15	36
	Building Blocks Library 8	6	17	21	25	69
	Building Blocks Library 9		1	2	8	11
	Cambridge English Readers Starter	11	26	36	86	159
	Cambridge English Readers 1	4	18	8	30	60
	Cambridge English Readers 2	1		4	6	11
	Cengage Page Turners 1	3	1	9	3	16
	Cengage Page Turners 2	8	21	23	28	80
	Cengage Page Turners 3	3	11	8	18	40

	Compass Young Learner's Classics 1			6	5	11
	Compass Young Learner's Classics 2			6	13	19
	Compass Young Learner's Classics 3		1	4	13	18
	Compass Young Learner's Classics 4			3	9	12
	Foundations Reading Library 1	2	3	3	10	18
	Foundations Reading Library 2	1		11	10	22
	Foundations Reading Library 3	6	1	10	21	38
	Foundations Reading Library 4	7	13	24	18	62
	Foundations Reading Library 5	6	13	22	26	67
	Foundations Reading Library 6	27	56	78	85	246
	Foundations Reading Library 7	35	72	78	89	274
	Helbling 1	1	3	10	12	26
	Macmillan Readers Beginner		3	20	23	46
	Macmillan Readers 2	14	38	43	81	176
	Macmillan Readers 3	7	5	6	8	26
	Oxford Bookworms 0	4	56	21	76	157
	Oxford Bookworms 1	49	57	55	111	272
	Oxford Bookworms 2	16	4	7	24	51
	Oxford Bookworms 3	3		5	10	18
	Oxford Classic Tales 1	1	1		8	10
	Oxford Classic Tales 3	2	7	8	4	21
	Oxford Classic Tales 4	4	19	13	10	46
	Oxford Classic Tales 5	7	30	21	22	80
	Oxford Dominos Starter	10	51	33	67	161
	Oxford Dominos 1	1	2	5	12	20
	Oxford Reading Tree 5	8	10	13	28	59
	Oxford Reading Tree 6	12	11	12	30	65
	Oxford Reading Tree 7	4	5	11	27	47
	Oxford Reading Tree 8	16	28	14	13	71
	Oxford Reading Tree 9	11	25	11	11	58
	Penguin Readers 0	1	8	6	25	40
	Penguin Readers 1	7	64	40	93	204
	Penguin Readers 2	56	73	60	96	285
	Penguin Readers 3	5	3	5	4	17
LRs	All Aboard Reading 2	3	14	12	6	35

Puffin Easy-to-Read 2 (Amber Brown)	1	9	5	8	23
Barrington	1	9	1	1	12
I Can Read 1	2	21	15	11	49
Magic Tree House		21	6	14	41
Nate the Great	10	42	37	24	113
Rainbow Magic		9	5	4	18
Ready-to-Read 2		4	4	12	20
Step into Reading 3	5	43	35	23	106
Step into Reading 4	7	41	47	21	116
Walker Stories	1	8	6	8	23

Table 1. Summary of six studies

Researcher(s)	Demography					ER			Writing		
	N	Country	Age/Year	# of hours	Amount	Kind of books	Place	Means for checking	Topic	Times	Scoring method
Hafiz & Tudor (1990)	25 (CG=24)	Pakistan	15 -16 years olds (secondary school)	90 hrs.	Not clear	GRs (Collins, Heinemann, Longman)	in class	by discussing	"My Family" "The Fire"	6 times (1st & 6th marked)	word count*, vocab (base* & ratio), syntactic maturity (T-unit length, sentence length, simple/ compound/ complex), accuracy of expression*(syntactically acceptable T-units, semantically acceptable T-units, spelling)
Lai (1993)	52	Hong Kong	11 -15 years old (secondary school)	24 hrs.	14.2 books (Mean)	GRs (Collins, Heinemann, Ladybird, Longman, Macmillan, Oxford, Witman)	in & outside	in: by discussing out: record sheet	"My Family"	2 times (pre- & post- tests)	word count*, error-free T unit count*, counts of simple/ compound/ complex sentences, types of error made (spelling*, tense, articles, verb agreement*, choice of vocabulary, pronouns, mechanics), holistic impression of the quality of writing
Tsang (1996)	48 (CG=96)	Hong Kong	12-17 years old	Not clear (24 wks.)	Not clear (assigned to read 8 books)	simplified classics original readers info-based books	outside	book reviews	"My Favorite Person"	2 times (pre- & post- tests)	content*, organization, vocab., lang. use*,mechanics, total score*
Lee &Hsu (2009)	43 (CG=43)	Taiwan	17-18 years old (vocation college)	25 hrs.	357.52 pages (Mean in 1st semester) 525.36 pages (Mean in 2nd semester)	GRs (Penguin 1-6, Oxford 1-6)	in class	record sheet	"The Moon Festival" "Summer Vacation"	2 times (pre- & post- tests)	content*, organization*, vocab.*, lang. use*, mechanics*, fluency*
Mermelstein (2015)	107 (CG=104)	Taiwan	3rd-year univ. students	7.25 hrs. (in class)	Not clear	GRs (Penguin 1-6, Oxford 1-6)	in & out	in: by observing out: record sheet	"Your Past Summer Vacation" "Your Future Summer Vacation"	2 times (pre- & post- tests)	content*, organization*, vocab.*†, lang use*†, mechanics*†, fluency*†
Park (2016)	28 (CG=28)	U.S.	1st year univ. 1st year graduate (ESL class)	16 wks. (8 hrs. in class)	Not clear	GRs (Oxford 2-6, Cambridge)	in & out	in: record sheet, discussion out: writing activity	"What is your attitude toward writing?"	2 times (pre- & post- tests)	content†, organization†, vocab.†, lang. use†

Note. CG = control group, hrs. = hours, wks = weeks, GRs = graded readers, vocab. = vocabulary, lang. = language, * = statistically significant, † = outperformed the control group