Vocabulary Gains in a Japanese University Extensive Reading Program

George Higginbotham
Eikei University of Hiroshima

Reference Data

Higginbotham, G. (2022) Vocabulary gains in a Japanese university extensive reading program. In P. Ferguson & R. Derrah (Eds.), *Reflections and new perspectives*. JALT. https://doi.org/10.37546/JALTPCP2021-16

Extensive Reading (ER) programs are established within many Japanese universities because of their numerous benefits (Nation & Waring, 2019). The issue for teachers is not whether to include ER, but how much they should require students to read and how to persuade institutions to invest in more books. In this pilot study into the relationship between ER and incidental vocabulary gains, these issues are explored through a discussion of data drawn from 160 undergraduates in a year-long English course requiring minimum reading goals. During the year, data were collected on reading rates (using MReader and XReading) and vocabulary knowledge (using Nation & Beglar's 2007 Vocabulary Size Test). The main finding was that a subgroup of consistent readers, who read over 40,000 words, had a significant positive correlation (r (26) = 0.41; p <0.05) between the amount read and VST gains. This implies that even a limited amount of ER over a one-year period can lead to measurable vocabulary gains.

多読（ER）プログラムは、その利点から日本の数多い大学に設置されている（Nation & Waring,2019）。そのような中、教師にとっての課題は、ERを取り入れるかどうかではなく、学生にどれだけ多読をさせるか、そして、より多くの本に触れさせるための書籍購入投資について教育機関をどう説得するかである。これらの問題を、最低限の読書目標を課している1年間の英語コースに所属する160名の学部生から収集したデータの考察を通して探った。1年間の読書量（MReaderとXReadingを使用）と語彙知識（Nation & Beglar's 2007 Vocabulary Size Testを使用）に関するデータを収集したところ、4万語以上を継続的に読んだ学生は、読書量とVSTの上昇に有意な正の相関（r (26) = 0.41; p <0.05）があることがにわかった。このことは、1年間の限られたERの導入であったとしても、測定可能な語彙の増加につながることを示唆している。

There is a growing consensus (e.g., Grabe & Stoller, 2019; Nation & Waring, 2019; Thomas, 2020; Webb, 2019) that Extensive Reading (ER) programs benefit foreign language (FL) learners in a variety of ways, including providing comprehensible and meaningful input, developing reading fluency, encouraging learner autonomy and increasing vocabulary knowledge. Accordingly, ER currently features as a component in many university English courses - particularly in Japan. The issue for many teachers now is not whether to include ER, but how much they should require their students to read. There is also a need to persuade institutions to invest in paper and/or online books by providing evidence that such an investment will yield measurable gains in language ability. While research has clearly demonstrated the benefits that ER has to motivation (Day & Bamford, 1998; Hougham, 2020; Yamashita, 2013), reading speed (Beglar & Hunt, 2014), reading fluency (Beglar et al., 2012; Samuels, 2006) and learner autonomy (Takase, 2012), fewer researchers have reported on incidental gains in vocabulary knowledge by FL learners in a sustained ER program. Of the research that has been done, this has mainly focused on learners who have read vast amounts (hundreds of thousands of words per semester), as opposed to the more modest amounts (tens of thousands of words) one might expect with lower proficiency learners or where ER is merely one component within a broader language course. The current study therefore represents a limited-exploratory study into the impact of ER on incidental gains in vocabulary knowledge.

Literature Review

In a series of classic studies of vocabulary retention (often referred to as the Clockwork Orange studies) Saragi et al. (1978) initially demonstrated that readers could acquire the meaning of words, in their L1, from the surrounding context of words that they could not have known previously. This finding was supported by a subsequent replication study with L2 learners a decade later (see Pitts et al., 1989) and then taken a step further by Horst et al. (1998). Horst concluded that acquisition rates for new words were not dramatic, although "...reading extensively will probably enrich their knowledge of the

words they already know, increase lexical access speeds, build network linkages between words, and more..." (p. 222). More recent investigations (Bordag & Rogahn, 2018; Waring & Takaki, 2003) within an EFL context also support extensive reading as a tool for consolidating vocabulary knowledge. While helpful in establishing that it is possible to retain a few words from reading a single book and also deepen knowledge of already known words, they did not offer an insight into how learners' mental lexicons increase in breadth over a sustained period of reading dozens of the short books that are typically used in modern ER programs.

Research has been done (Aka, 2019; Elley & Mangubhai, 1981; Lee, 2007; Nishizawa et al., 2010; Nishizawa et al., 2017) that specifically explored vocabulary gains that an ER program can bring to FL learners, although as will be argued, problems with methodology did not allow strong claims to be made. The first of these (Elley & Mangubhai, 1981) was a study that justified a pioneering program and established ER as a methodology that could lead to significant language gains. Unfortunately, this study was run at a time when standard tests of vocabulary had yet to be developed and the learners did not have access to graded readers written specifically for them using high-frequency vocabulary and simplified grammar. It is therefore difficult to generalise their findings to modern programs, which due to the more appropriate graded readers would be expected to be more effective.

An early study, of what might be termed a modern ER program, compared the vocabulary gains of 108 Taiwanese FL learners in a course with an ER component to a control group (n=137) without an ER component (Lee, 2007). The ER groups, as measured by a pre- and post-course Vocabulary Levels Test (Schmitt, 2000), significantly outperformed the control group. One issue with this study was that a version of Nation's VLT (a diagnostic test) was used to measure gains in vocabulary proficiency. Since Lee's study, other tests of receptive vocabulary, such as the Vocabulary Size Test (Nation & Beglar, 2007) have been designed and validated (Beglar, 2010) to better capture gains in general vocabulary proficiency. Another criticism of Lee's study was that the amount of reading learners did over the year was not recorded, which is unhelpful if we want to advise learners on how much they should read.

Following Lee (2007), researchers found significant correlations between the number of words read and scores by Japanese learners on TOEIC tests (Nishizawa et al., 2010). Although widely cited, they did not offer compelling evidence due to the small group sizes (n = 9; n =13; n = 8). A follow-up study over a 6-year period (Nishizawa et al., 2017) examined similarly small group sizes. Another shortcoming was the researchers did not measure student TOEIC scores prior to the ER study, but rather compared average TOEIC scores, meaning that the amount each learner gained during the study period was not determined.

One recent study utilized large group sizes (experimental group 200; control group 205) and found that over a year the first-year high school learners made significant gains in their vocabulary knowledge (Aka, 2019). While Aka estimated that learners read 115,000 words (on average), as with Lee (2007) the precise number of words that each student read were not measured. We might also note that in Aka's (2019) study the vocabulary scores were part of a larger test, with only 16 items being used to judge vocabulary knowledge. It is questionable whether such a small number of test items could reliably measure gains in vocabulary. While useful, these studies indicate that we are at an early stage in understanding how much ER is required for L2 learners to make measurable vocabulary gains. As well as building on this research, the current study responds to the call by Webb (2019) to "...focus more on longitudinal studies of incidental vocabulary learning through exposure to multiple texts" (p. 234).

In evaluating the effect of reading on vocabulary acquisition we can divide the literature into two distinct approaches. Some researchers take what might be described as an *ant's eye* perspective, while others take a *bird's eye* perspective. The ant's eye approach (Horst et al.,1998; Waring & Takaki, 2003; Webb & Chang, 2015) carefully examined multiple aspects of word knowledge for a pre-selected (and limited) number of words within one (or a few) text(s) over a short time-frame with a small sample of learners. The bird's eye approach (Aka, 2019; Elley & Mangubhai, 1981; Lee, 2007, Nishizawa et al., 2010; Nishizawa et al., 2017) took a broader view and examined the effects of ER on a language program over a longer time-frame, usually with just one measure of language ability over a year. Both perspectives offer complementary insights, and can both be used by teachers aiming to persuade their institution to adopt ER within their English program, or perhaps to convince their library to expand the graded-reading section.

Research Questions

Mindful of problems with prior studies (Aka, 2019; Lee, 2007; Nishizawa et al., 2010), I decided to collect data on the precise number of words read by a sample of FL learners and measure their vocabulary gains by way of an 80-item test of receptive vocabulary knowledge both at the start and end of a one-year course. Based on the previous studies findings, the expectation was that the learners who read the most would improve their vocabulary scores the most. It was also expected that an analysis of students who read to

their full potential, in a consistent way (rather than just cramming their reading into the last few weeks of a semester), would yield more accurate data. Furthermore, by examining the number of words read in that period it was anticipated that minimal thresholds might be identified, leading to recommendations on reading goals for FL learners at this level. The current study was therefore designed to answer the following questions:

RQ1. Is there a relationship between the amount FL learners read in a one-year ER program and gains in incidental vocabulary knowledge?

RQ2. Is there a statistical difference between consistent and inconsistent readers?

RQ3. What is the minimum amount of reading required for FL learners to make statistically significant gains in a vocabulary test?

Methodology

Participants

Following institutional approval, data on reading rates and vocabulary knowledge were collected from 160 female L2 learners at a Japanese university between May 2020 and February 2021. All participants were required to complete an ER component as part of a compulsory Basic English Course. Although over 200 undergraduate students were available at the start of the study, not all completed the tests or agreed to their data being used, leaving a final participant number of 160 that had given their consent. Participants were of widely varying levels of English ability. VST scores indicated that the lowest levels were struggling with the most frequent 1,000 words, and that those at the highest level had a receptive understanding of around 7,000 words. This diverse spectrum of abilities reflected the mix of learners from four different faculties: Nutrition, Childcare, Design, and English. In contrast to Day and Bamford's principle that "reading is its own reward" (2002, p. 138), the course teachers took the position that in order to compete with the demands of other university courses, some form of *push* was needed: reading goals, short tests of comprehension and a reading contest were therefore included as part of the ER experience. Within their compulsory 4-skills Basic English course, non-English-majors were required to read a minimum of 20,000 words over the year, and English-majors were required to read 40,000 words to get full credit for the ER component (25% of their final grade). To encourage reading beyond these minimum requirements, a university-wide reading contest was run that gave a range of prizes at the end of each semester (see Hougham, 2020 for details) dependent on how much was read. Many of the students were therefore motivated to read well beyond their class requirements, with the top prize being awarded to the first student to read beyond 300,000 words in a semester.

Measuring reading rates

To measure reading rates, online reading management systems (MReader and XReading) were used. The MReader website (Robb, 2015) is designed to assist extensive reading programs and allows teachers to keep track of how many words each learner has read over a period of study. The XReading system is similar in that it keeps a running total of the number of words read, although it also allows learners to read the material online. Both systems check that books have been understood through a comprehension quiz, with students needing to score over 60% in order to be credited with the words for that book. Each system keeps a cumulative record of the number of words each learner has read for these *passed* books. And as each book that is read is time-stamped, this also allows for an objective identification of those students who are reading in a *consistent* way (which we defined as reading for at least 13-weeks of each 15-week semester). The explanation of these online systems and the quizzes was in the learners L1 (Japanese). The books available to the learners were suitable in terms of length, vocabulary and grammatical complexity. At the easier end (e.g., Oxford Reading Tree Level 3) these books contain fewer than 100 words per title, and at the more difficult end (e.g., Cengage Page Turner series level 7) contain over 10,000 words per title. Participants were able to borrow paper books from the university library, directly from the teachers in their classes, and also use the online reading site (XReading.com). It ought to be noted that although XReading is a commercial site it complements MReader in that they both draw from the same set of quizzes and both collect data on the number of words read in the same way. In the spirit of ER, which emphasizes choice, learners were not limited with the medium they read: paper books, digital books, or a mix of both digital and paper books. The participants selected the kind of books they read, although they were guided to read appropriate levels by their teacher. Some class time was allocated to reading but, as it competed with other aspects of these 4-skills classes, much of the reading was considered homework. An unexpected factor that necessitated all students using just the digital books, for at least part of each semester, were campus/library restrictions imposed due to the Covid19 pandemic.

Measuring vocabulary gains

To measure vocabulary gains, participants took an abridged bilingual version of the Vocabulary Size Test (VST) (Nation & Beglar, 2007) at the start of the course in May 2020 and then again in February 2021. The tests were in a digital format and delivered via Google Forms during class times, which most completed within 20 minutes using their smartphones. The explanations for these tests were in the participants' L1 (Japanese).

The VST is a multiple-choice test of receptive vocabulary, is freely available to teachers and has been thoroughly validated in both its original (Beglar, 2010) and bilingual formats (Nguyen & Nation, 2011). In order to limit guessing an *I don't know* option was added (following Zhang, 2013) and the full test (140 items) was shortened to 80 items following Nguyen and Nation (2011) and McLean et al., (2014) by removing items that were at frequency levels far beyond the learners' abilities. Following McLean et al. (2014) the 80 items comprised of 10 items selected from each 1,000-word frequency level up to the 8,000-word level. While the expectation was that the books being read would generally contain words within the most frequent 3000 words, as students were of a broad range of abilities and were selecting from a wide range of graded readers (that also contain many lower frequency items) it was decided to use a test that went beyond just the high-frequency band.

As well as being easy to administer to a large group (and having been thoroughly validated with L2 learners) there are good reasons to use the VST (see Nation & Beglar, 2007 for a detailed discussion). One reason is that it gives credit for partial knowledge of a word's meaning and therefore allows learners to demonstrate gains on the edge of their mental lexicons: words understood at a very basic level but perhaps not yet available for production. Another reason is that it taps the receptive aspect of vocabulary knowledge, matching the receptive nature of reading. The scores on the two VSTs were compiled into a spreadsheet alongside the number of words read in the year. A Pearson correlation was calculated between the amount of vocabulary gained (VST2 minus VST1) and the number of words read.

Results

As can be seen in Table 1, the learners made a gain of around 5 points (on average) on the pre- and post-vocabulary tests: a gain of around 500 words. While encouraging, it should also be noted that there was a considerable amount of deviation from this mean value. In the pre-course VST, for example, the lowest score was 9 and the highest 77 from a possible maximum of 80 points. The gains on these pre- and post-tests also fluctuated considerably, with some learners making negligible (or negative) gains and others making large gains (e.g., +20 points). Table 1 shows the greatest gains were made in the top 2 frequency bands, smaller gains in the lower frequency bands and negligible gains in the lowest frequency band. This steady decrease in gains as the words tested became less frequent is unsurprising; as students would have had many more opportunities to meet (and so learn) the higher frequency bands words while reading.

Table 1

Pre/post VST Scores at each 1000 Word Band

		1k	2K	3K	4K	5K	6K	7K	8K
VST1	AV	83.52	64.92	58.36	62.08	57.41	50.82	47.35	52.63
	STD	16.77	20.45	19.02	20.41	19.69	24.79	22.95	23.20
VST2	AV	91.15	71.15	63.95	67.64	62.77	55.57	51.74	54.73
	STD	10.09	21.96	17.44	21.28	18.94	23.04	23.58	22.69
Gain (VST2 - VST1)		7.63	6.23	5.59	5.56	5.36	4.75	4.39	2.10
								Average Gain	**5.20**

Note: (n = 160)

The wide range of gains on the VST tests were expected and reflect the ability levels of the Non-English and English-major participants. As was also expected, there was a wide range in the number of words that students read in the ER program: typically, the English-majors read far more. Table 2 presents the amount of words read by major. It is reasonable to assume that the differing amounts of reading were partly due to ability and levels of motivation.

Table 2

Average Number of Words Read

	Semester 1	Semester 2	Year (Sem1+Sem2)
All (n =160)	17,921	25,712	43,472
Non-English-majors	16,671	23,072	39,743
English-majors	27,784	47,763	72,893

Directly responding to RQ1, for the whole group (n = 160) there was no significant correlation between the amount of words read and the gains in the VST scores. Viewing the entire cohort however conflates consistent with inconsistent readers. Many learners merely satisfied their class requirement (20,000 words/yr for Non-English-majors; 40,000/yr for

English-majors) and then stopped reading a few weeks before the end of semester. Others did not read at the start of the semester and then crammed the required reading in the last few weeks. In order to focus the analysis on consistent readers (RQ2) who were reading close to their potential, a subgroup of learners was selected who had read over double their class goals and for at least 13 of each 15-week semester (see Table 3).

Table 3

Average Pre/post VST Scores for the Consistent Readers (n = 26)

VST1 score	STDEV	VST2 score	STDEV	Gain (VST2 - VST1)	STDEV
55/80	12.2	58/80	11.9	3	5.2

While the average gain was just 3 points in the pre-post VST tests, with these *consistent readers* there was a positive and statistically significantly correlation ($r = 0.41$, $p < .05$) between the number of words they read and the gains they made on the VST. Based on the effect size (r^2), 17% of the variance of the VST scores can be said to be shared by the number of words read. Within this *consistent reader* group, for example, the learner who read the most (534,816 words in a year) made a 12-point gain in her VST score. In this same group, the learner who read the least (40,609 words) dropped 2 points on her second test. A reanalysis of the subgroup with 4 mild outliers removed also gave a similar positive and statistically significantly correlation, the mild outliers were therefore retained. There were no extreme outliers (see Appendix).

These results also inform RQ3 in that the class goals of 10,000 words per semester for Non-English majors and 20,000 words per semester for English-majors were generally insufficient to affect vocabulary knowledge, as measured by the VST. However, when learners read over double these amounts (>20,000 words or >40,000 words respectively) then gains in vocabulary could be detected by the pre- and post-course VST measure. The average amount read by the *consistent readers* (n=26) was 118,579 words in a year, which ranged from 40,609 to 534,000 words. Another way to view this would be that for learners who read over 40,000 words in a year (>20,000 per semester), a statistically significant relationship between the amount they read and gains in vocabulary knowledge could be discerned.

Discussion

With this exploratory study, which took a bird's eye perspective, measurable gains could be discerned when learners read in a consistent way over 40,000 words in a year. The gains made on the VST for the *consistent* group correlated significantly with the number of words each learner read. While the size of this group is not large enough to make strong claims (n =26) this finding suggests that even a modest amount of reading within a general English course can lead to discernible gains in vocabulary. This is important as previous studies of this kind (e.g., Aka, 2019; Nishizawa et al., 2010, Nishizawa et al., 2017) focused on courses where students had achieved far greater amounts of reading. We cannot however interpret this to mean that reading below a 40,000-word threshold will not lead to vocabulary gains. This is because the gains measured by the VST inevitably underestimate the amount of word learning that is occurring. It treats word knowledge as a simple known/unknown dichotomy that does not accommodate the widely accepted idea (Schmitt, 2000; Webb, 2019) that word knowledge is gradually accumulated over multiple exposures. An item like *shoe*, for example, was answered correctly on both the pre- and post-course tests by many of the learners, which might mean that no extra knowledge was gained for this word. It is however just as likely that on the first attempt the learner knew the L1 equivalent (but perhaps not much else), and then after meeting the word numerous times in the books began to understand that a word like *shoe* is often used with an *s* to make the plural form (as most people have two feet) and is also often used with verb phrases such as *put-on* or *take-off*. After more encounters it might also occur to the learner that this word is a noun and is often preceded by color adjectives (*brown, black*). This gradual layering of knowledge over the year would not have been picked up by the VST, which views word knowledge as a known/unknown dichotomy. The implication is that while significant gains in word knowledge were not detected by the VST for learners who read less than 40,000 words in a year this does not mean that word learning did not occur. On the contrary, it is probable that a gradual accumulation of different aspects of vocabulary knowledge did occur at each encounter.

So, while we can make tentative claims (based on the limited data presented here) that word knowledge was gained for those learners who read more than 40,000 words per year, we cannot confidently state a reading amount below which ER is ineffective (due to the limitations of the type of test used). To identify this threshold reading rate, below which no vocabulary learning is occurring, would probably require an approach closer to the ant's eye approach discussed earlier, with multiple measures of vocabulary knowledge. An alternative to the VST, the Generalized Vocabulary Test (Suk, 2021) has

recently been developed in a Korean context that might offer a more accurate method of measuring vocabulary knowledge acquired through ER. However, being in essence just another kind of test of receptive vocabulary knowledge, it may well suffer from the same limitations as the VST. Despite this, future studies in measuring gains in incidental vocabulary knowledge through ER should be encouraged to compare alternative measurement tools, such as Suk's GVT, with the more established tools such as the VST. Future studies might also be able to address another unavoidable limitation of this study, the problem that within the course the learners were not exclusively involved in ER but (as this was a 4-skills general English course) also engaged in a range of other language learning activities. It is therefore conceivable that the observed vocabulary gains by the *consistent reader* group merely identified those who were well motivated (and so worked hard) at all areas of their English. To address this issue a longitudinal study of a course devoted to reading would be welcomed.

Conclusions

For those who already see ER as beneficial to L2 learners for a host of other reasons, the correlation found in this study between the number of words read and gains on a pre/post-course vocabulary test is yet another positive sign. Given the limitations noted in the discussion section the main research question can however only be tentatively affirmed. For the *consistent readers* in this study, a gain in vocabulary knowledge was observed that increased in a statistically significant proportion to the amount they read. This study therefore gives a little more ammunition to those who are trying to persuade their institutions to loosen their purse strings and invest in paper books and/or online reading. While unable to add much clarity to the third research question (the minimum number of words L2 learners need to read in a year to make measurable gains), the finding that there is a correlation between those that read over 40,000 words in a year and their gains in vocabulary knowledge means that we have now edged a little closer to this threshold. This amount is considerably less than reported in Aka's (2019) study that found vocabulary gains with learners who were reading around 100,000 words a year. While this study offers evidence to support the claim that FL learners reading over 40,000 words per year in an ER program can expect to make gains in their vocabulary, this ought not to be viewed as a lower limit. This study can neither deny nor support the claim (due to the limitations of the vocabulary measure used) that this is a threshold below which vocabulary gains would not be expected. Further research using multiple measures of vocabulary, and alternative test types, will be required to identify this lower limit with greater precision.

The main take-away from this study is that ER is not just about increasing motivation or reading speed. Foreign language learners who make a consistent and sustained effort with ER can also expect to make measurable gains in their vocabulary knowledge.

Bio Data

Following his Ph.D. with Swansea University, **George Higginbotham** has pursued a research agenda in the vocabulary development of L2 learners. As an EFL educator in Japan and the UK for over 20 years, his research interests include: measuring productive vocabulary, word association testing, and evaluating programs for extensive reading and academic writing. George is affiliated with Eikei University of Hiroshima and can be contacted at <gmhigginbotham@gmail.com>

References

Aka, N. (2019). Reading performance of Japanese high school learners following a one-year extensive reading program. *Reading in a Foreign Language, 31*(1), 1-18. https://doi.org/10125/66747

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 21,* 101–118. https://doi.org/10.1177/0265532209340194

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language, 26*, 29–48. https://doi.org/10125/66684

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

Bordag, D., & Rogahn, M. (2018). The role of literariness in second language incidental vocabulary acquisition. *Applied Psycholinguistics, 40*(2), 399–425. https://doi.org/10.1017/S0142716418000620

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

Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a foreign language, 14*, 136–141. https://doi.org/10125/66761

Elley, W. B., & Mangubhai, F. (1981). *The impact of a book flood in Fiji primary schools*. New Zealand Council for Educational Research.

Grabe, W., & Stoller, F. L. (2019). *Teaching and researching reading*. Routledge.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*, 207–223. https://doi.org/10125/66953

Hougham, D. (2020) Investigating the effect of MReader quizzes on Japanese university students' attitudes toward reading in English. *Extensive Reading World Congress Proceedings, 5,* 137-155.

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

McLean, S., Hogg, N., & Kramer, B. (2014). Estimations of Japanese university learners' English vocabulary sizes using the vocabulary size test. *Vocabulary Learning and Instruction, 3*(2), 47-55. http://dx.doi.org/10.7820/vli.v01.1.2187-2759

Nation, I.S.P. and Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13. ISSN: 0289-7938

Nation, I. S. P., & Waring, R. (2019). *Teaching extensive reading in another language*. Routledge.

Nguyen, L.T.C., & Nation, P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal, 42*(1), 86-99. https://doi.org/10.1177/0033688210390264

Nishizawa, H., Yoshioka, T., & Fukada, M. (2010). The impact of a 4-year extensive reading program. In A. M. Stoke (Ed.), *JALT2009 Conference Proceedings.* Tokyo: JALT.

Nishizawa, H., Yoshioka, T., & Ichikawa, Y. (2017). Effect of a six-year long extensive reading program for reluctant learners of English. *Modern Journal of Language Teaching Methods, 7*(8), 116-123.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign language, 5*(2), 271-275. https://doi.org/10125/67017

Robb, T. (2015). Quizzes – A sin against the sixth commandment? In defense of MReader. *Reading in a Foreign Language, 27*(1), 146-151. https://doi.org/10125/66706

Samuels, S. J. (2006). Reading fluency: Its past, present, and future. *Fluency instruction: Research-based best practices, 2*, 3-16.

Saragi, T., Nation, I. S. P., & Meister, F. (1978). Vocabulary learning and reading. *System, 6*, 72–78. https://doi.org/10.1016/0346-251X(78)90027-1

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge University Press.

Suk, N. (2021). Developing a sensitive but generalizable measurement of vocabulary gains from self-selected extensive reading. *System, 101*, https://doi.org/10.1016/j.system.2021.102614

Takase, A. (2012). The effectiveness of sustained silent reading in becoming autonomous learners. *Gaikokugo Kyoiku Forum [Forum for Foreign Language Education] 11*, 1-14.

Thomas, N., (2020). Incidental L2 vocabulary learning: Recent developments and implications for future research. *Reading in a Foreign Language*, *32*(1),49-50. https://doi.org/10125/66576

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*, 130-163.

Webb, S. (Ed.). (2019). *The Routledge Handbook of Vocabulary Studies.* Routledge.

Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667-686. https://doi.org/10.1177/1362168814559800

XReading [website]. Available at: https://xreading.com/

Yamashita, J. (2013). Effects of extensive reading on reading attitudes in a foreign language. *Reading in a Foreign Language, 25*(2), 249-263. DOI:10125/66872

Zhang, X. (2013). The I don't know option in the Vocabulary Size Test. *TESOL Quarterly, 47*(4), 790-811. https://doi.org/10.1002/tesq.98

Appendix

Gains Made by Consistent Readers

Student	VST GAINS (VST2 – VST1)	Total words read in 2 semesters
1	-2	40609
2	2	40857
3	-8*	41307
4	-2	42674
5	8	43160
6	-3	43937
7	-6*	47370
8	0	52357
9	-5*	56811
10	7	57602
11	7	60133
12	8	60267
13	6	63531
14	0	69308
15	3	74671

Student	VST GAINS (VST2 – VST1)	Total words read in 2 semesters
16	11*	93769
17	3	102862
18	5	108878
19	5	110380
20	4	111720
21	3	113806
22	9	191404
23	3	238635
24	6	261973
25	2	420218
26	12*	534816

* = mild outliers