Extensive Reading and TOEFL ITP Scores

Marcel Van Amelsvoort

Juntendo University

The TOEFL ITP results and extensive reading (ER) progress of two cohorts of first year university students in Japan (n = 400) were examined to look for if and how ER contributes to score improvements on the test. The study tracked student ER behavior over 27 weeks—reading volume, time spent reading, amount of listening while reading, reading rate, and book level—and compared them to TOEFL ITP scores on an April pretest and December post-test. Results showed that reading volume and reading time had a small but statistically-significant effect, while reading rate and book level did not. Listening while reading showed a small, statistically-significant effect, but the nature of the data did not allow that to be stated with confidence. Implications for curriculum design are discussed.

The TOEIC test and to a lesser extent the TOEFL ITP test are widely used to gauge proficiency improvements and to measure English language program effectiveness in Japan and several other countries. According to the ETS Japan website, as of 2019, a total of 178 universities across Japan were making regular use of the TOEFL ITP. Students, as well as potential employers also use test achievement to ascertain proficiency, and a wide range of materials are produced by publishers specifically for learners hoping for higher scores. A quick search on Amazon Japan returns over 10,000 titles for the more popular TOEIC test, and over 300 for the TOEFL ITP.

In EFL settings, one relatively common practice to maximize language exposure and engagement and thus improve learning outcomes is to include an extensive reading (ER) component in a program, particularly additive ER, where most of the reading is done out of class (Robb & Kanno, 2013). Extensive reading, which provides large amounts of language input with a focus on meaning, affords opportunities for incidental learning in an enjoyable and motivating activity (Day, et al., 1991; Elley, 1991). It has been shown to improve many aspects of student proficiency (Nakanishi (2014), especially reading speeds (McLean, & Rouault, 2017) and vocabulary (Suk, 2017). The current study looks at one program that features ER as a regular compulsory activity embedded in a unified content and language integrated learning (CLIL) liberal arts syllabus at a private university in Japan. The study sought to discern whether a year-long ER program had any significant effect on student TOEFL ITP achievement.

Literature Review

Studies focusing on the effects of ER on test performance are not uncommon, but as the learning takes place over long periods of time, it is difficult to control for other learning experiences which of course also affect scores on standardized tests.

Early studies looked for correlations between different learning behaviors and TOEFL success. Gradman & Hanania (1991) found through surveying a small number of international students that out of class reading contributed most to success on the test. Constantino, Lee, Cho, & Krashen (1997) in another study of international students (n = 43) found moderate correlations between TOEFL success and a composite measure of out-of-class reading, calculated with scores for free reading (r = .39) and the number of books read (r = .447). The authors observed that the better TOEFL scorers were those students more likely

to read in the L2. While neither of these studies focuses on extensive reading, they helped to popularize the idea of reading one's way to TOEFL success in a mostly enjoyable manner.

Studies on ER and TOEFL ITP scores are rare, and are not very optimistic. Williams (2009) looked at the relationship between ER and TOEFL ITP score improvements with a large number of university students at a university in Japan (n= 564) but only over the short time span of one semester. The number of words read was not tracked in the study, but the number of books (5.2 books per student on average) and pages (178 pages on average for the less proficient students and 228 pages on average for the more proficient students) was. The control group of 205 students did no extensive reading at all. It turned out that these students had greater gains on the TOEFL ITP test on average than the students who did the ER. This was true at all levels of proficiency: high (11.4 for non-readers and 8.9 for readers), middle (10.9 for non-readers and 6.9 for readers), and low (24.4 for non-readers and 19.4 for readers). Although Williams found that many (but not all—some experienced score drops) of the most prolific readers (350 or more pages) showed greater than average gains for their respective level groups, he ascribed the disappointing results of the ER program on average to the insufficient quantity of reading in general and the low level and types of the books students were reading (fiction in this study). He also mentioned that ER alone would likely be insufficient to produce large gains without dedicated study of listening, grammar, and other skills and strategies associated with the test. Van Amelsvoort (2019) looked at TOEFL ITP score changes in a program with almost two years of ER. One cohort ($n = 105$) read an average of 314,765 words during that time. An analysis of the data found no correlation with raw ER reading (word counts) and TOEFL ITP scores. When book level was factored in, however, a small correlation was observed with reading points (word count × reading level) and the total score (.24) and the reading section (.26).

Several studies which examined the relationship between ER and TOEIC improvements are also worth mentioning here as they involve issues of reading volume, control of language exposure and learning experiences, and a strong emphasis on standardized test performance as a measure of student progress. Nishizawa, Yoshioka, & Fukada (2010) and Nishizawa, & Yoshioka (2016) looked at longitudinal data for student reading amounts, levels, and TOEIC scores. They found that, on average, learners need to read 350,000 words to start to see improvement on the TOEIC test, and 628,000 words to exceed a score of 400. Great variability among learners was found, however. These studies emphasized the importance of reading at great volume and managing reading with slow incremental increases in book length and level over several years. Other researchers working with shorter durations for their interventions have not found TOEIC scores easily moved by ER. O'Neill (2012) found that additional gains made through student participation in an ER program were not reflected in student TOEIC reading section scores. Likewise, Storey, Gibson, and Williamson (2006) and Rutson-Griffiths & Rutson-Griffiths (2018) found a very small effect for students doing ER, the former when TOEIC reading questions were also assigned and the latter when total words read and quiz scores achieved were considered. Carney (2016) found "almost no statistically significant relationship between increased ER reading and improvement in TOEIC reading scores" (p. 69). Lyddon & Kramer (2019) found a positive correlation between word count and TOEIC scores but wondered whether ER reading might indicate general compliance rather than being a direct cause of TOEIC reading score improvements. And finally, in a study illustrating the importance of multiple fluency-developing activities, Milliner (2021) combined ER with timed reading and repeated oral reading and found that

learners who did more ER had significantly greater gains in reading rate and improved both listening and reading scores on the TOEIC.

Another aspect to consider is the use of audio and its effect on scores. In recent years, advances in technology have made it enormously easier for students to listen to an audio narration of the books they are reading extensively. This is particularly true for online ER platforms such as the Xreading Virtual Library (xreading.com), which not only facilitates listening while reading, but tracks the time of use. Despite the potential of listening, there does not seem to be conclusive data showing an effect on TOEFL ITP scores. In addition to reading, Van Amelsvoort (2019) also looked at the effect of listening. However, the results were not conclusive because most students did not do any listening while reading at all, and the students who did do so, did so with amounts so small that it was impossible to say anything about the possible effect of listening. Gobel (2011), on the other hand, found a positive effect for ER with listening with 162 university students in Japan, calling it a "solid predictor" (p. 45) of TOEFL ITP improvement. He found that ER with listening accounted for 30% in score changes between pre- and post-TOEFL ITP scores, and the reading level accounted for 18% of the variance. That is, students who read higher level materials (A2 level books versus A1 for the other groups) experienced greater gains. His study had suspiciously small word count targets for many of the students (10,000-100,000) compared to most ER studies, however, and students were concurrently taking three other required English courses, the content of which was not described in the study.

Despite the proven effectiveness of ER with improving reading rates and vocabulary, frustration with finding a strong, clear connection between reading for fluency and score improvement on standardized tests is common (Carney, 2016, Lyddon & Kramer, 2019, Rutson-Griffiths & Rutson-Griffiths, 2018, Storey, Gibson, and Williamson, 2006, and Van Amelsvoort, 2019). Carney (2016) gave voice to this notion when he raised questions about quantity and quality in ER programs. He wondered whether it were possible for university students to read enough in a year or two to see significant improvement in scores, and whether a well-designed ER program might lead to better scores when a poorly-designed one does not.

The Current Study

There is a dearth of studies on ER and TOEFL ITP scores given the large number of schools using it to monitor progress of student proficiency gains. Considering the interest in CLIL and EMI (English Medium Instruction) courses in recent years, the TOEFL ITP, as a reliable and affordable measurement tool, will likely continue to be widely used. For many of these schools, ER is a mandatory and graded feature of their program.

Few studies to date have examined programs with large numbers following a unified curriculum. To my knowledge, studies have not attempted to control data by removing students who have participated in study abroad programs, or attempted to control for student dishonesty. Few studies have looked at various aspects of ER performance, including both time on task and word counts for ER, reading rates, extensive listening, and comprehension quiz results.

For the current study, effort was made to avoid some of the problems of research with ER, namely controlling the exposure to English within and beyond the syllabus, and obtaining accurate data of participant engagement. The institution where the study took place featured a unified syllabus with four 90-minute classes per week, all of which used identical materials. The data was collected from two cohorts over two years. In addition, the collection of data was timed to give the most accurate record of actual student reading and listening while reading possible. One issue

in programs making use of mandatory ER is the amount of academic dishonesty that takes place. Students regularly claim to have read books which they actually have not read, for the purpose of achieving the necessary word counts to receive points for the ER assignment. Anecdotal evidence has shown that the vast majority of this falsification occurs in the week or two before the assignment closes. For that reason, the ER data was collected on the same day that the TOEFL test was administered, almost one month before the ER deadline.

The current study had the following research questions:

Q1: To what extent did the number of ER words read affect changes in TOEFL ITP scores?

Q2: To what extent did the amount of time spent reading extensively affect changes in TOEFL ITP scores?

Q3: To what extent did the amount of time spent listening while reading extensively affect changes in TOEFL ITP scores?

Q4: To what extent did the attained reading speed affect changes in TOEFL ITP scores?

Q5: To what extent did the average graded reader book level read by learners affect changes in TOEFL ITP scores?

Q6: Did the top readers (by volume) or the top test score improvers show a connection between ER and test scores?

Methodology

Participants

The male and female participants were two cohorts of first year students in an international liberal arts program at a private university in Tokyo that accepts 240 students per year. Students were enrolled in a compulsory course that featured an ER component worth 10% of their grades. The program made use of a unified syllabus, thus all students experienced the same number of contact hours and with mostly the same materials and activities. A number of students were excluded from the study for a variety of reasons. Some were repeating the course due to a previous failure, some had gone abroad to study over the summer, some did not engage in any ER, and some did not take either the pre- or post-test.

Instruments

Students accessed, read books, and had their performance tracked on the Xreading Virtual Library (xreading.com) to which they were all given subscriptions. Upon matriculation into the program in April, a TOEFL ITP test was administered. This functioned as a pre-test. ER was assigned for both the first and second terms, with the second TOEFL ITP taking place at the end of December, three weeks before the end of the term. Students were also given a small summer ER assignment. Thus, the majority of students were engaged with ER for approximately 27 weeks.

Procedures

The April TOEFL ITP was used to determine basic student proficiency and to stream students into one of three levels. Students began ER in April with weekly assignments and monthly benchmarks set by individual teachers. Word count targets for each of the 16 classes depended on the level of the class, with upper level classes expected to read 100,000 words per term, intermediate level students expected to read 80,000 words per term, and lower level students expected to read 70,000 words per term. All students were encouraged to read beyond these targets, however, and were given bonus points for doing so. Students were shown how to access the audio for books and were regularly reminded by instructors to turn on the audio. The benefits of reading while listening were explained to students.

On the same day of the post-test TOEFL ITP, data for ER performance was collected. This date was about three weeks before the official closing of the ER program. It was chosen to avoid counting any reading done after the test. Furthermore, due to anecdotal evidence of academic dishonesty regarding the reporting of reading (cheating with the online system) that typically occurs just prior to the ER program deadline, having the post-test well before the deadline for the ER program helped to get a more accurate picture of student ER performance. Data for ER performance and TOEFL ITP achievement were then compared.

Results

The descriptive statistics for student performance are shown in Table 1. As can be seen, there are large variations in the numbers, something unsurprising given the large number of students.

Linear regression analyses

Linear regression was used to measure the impact of the various factors on the change in student TOEFL ITP scores between April before the beginning of the ER intervention and December, near the end. Data is reported here according to the research questions.

Q1: To what extent did the number of ER words read affect changes in TOEFL ITP scores?

A linear regression was conducted to examine how well the number of words read could predict improvements in TOEFL ITP scores. A scatter plot showed that the relationship between words read and score improvement was positive and linear and did not reveal any bivariate outliers. An analysis of standard residuals showed that the data contained no outliers (Std. Residual Min. = -2.93, Std. Residual Max. = 2.99). Independence of residual errors was confirmed with a Durbin-Watson test (d = 1.875). Residual plots showed homoscedasticity and normality of the residuals. The number of words read significantly predicted score increases, $F(1, 398) = 20.78$, $p < 0.001$, accounting for 5% of the variability in score changes with adjusted R2 = 4.7%. This is a weak predictive relationship (Cohen, 1988). The correlation between number of words read and score improvement was statistically significant, $r(398) = .223$, $p < .001$. While the results were significant, the effect was quite small.

Q2: To what extent did the amount of time spent reading extensively affect changes in TOEFL ITP scores?

Table 1. Descriptive statistics

	Valid	Missing	Mean	Std. Deviation	Min.	Max.
Word count	400	0	107,892	59,730	85	491,154
Reading time (mins.)	400	0	993.50	577.25	4.00	4,038
Listening time (mins.)	400	0	245.31	369.46	0.00	2,428
Reading speed	400	0	116.27	38.06	32.00	400.60
Avg. book level*	400	0	5.4	1.5	1	9

Note. * Book level is per the Extensive Reading Foundation Grading Scale (n.d.), with Level 1 =CEFR pre-A1 level and Level 9 = CEFR B1 level.

A linear regression was conducted to examine how well the amount of time spent on reading could predict improvements in TOEFL ITP scores. A scatter plot showed that the relationship between reading time and score improvement was positive and linear and did not reveal any bivariate outliers. An analysis of standard residuals showed that the data contained no outliers (Std. Residual Min. = -2.98, Std. Residual Max. = 3.11). Independence of residual errors was confirmed with a Durbin-Watson test (d = 1.862). Residual plots showed homoscedasticity and normality of the residuals. The amount of time spent reading significantly predicted score increases, F (1, 398) = 27.5, $p < .001$, accounting for 6.5% of the variability in score changes with adjusted R2 = 6.2%. This is a weak predictive relationship (Cohen, 1988). The correlation between amount of time reading and score improvement was statistically significant, $r(398) = .254$, $p < .001$. This result, while still relatively weak, emerged as the strongest predictor of variance.

Q3: To what extent did the amount of time spent listening while reading extensively affect changes in TOEFL ITP scores?

A linear regression was conducted to examine how well the amount of listening while reading could predict improvements in TOEFL ITP scores. A scatter plot showed that the relationship between listening time and score improvement was positive and linear and did not reveal any bivariate outliers. An analysis of standard residuals showed that the data contained no outliers (Std. Residual Min. = -2.97, Std. Residual Max. = 2.77). Independence of residual errors was confirmed with a Durbin-Watson test ($d = 1.780$). Residual plots, however, failed to show homoscedasticity due to the large number of students who did no listening while reading at all. The amount of time students spent listening significantly predicted score increases, F (1, 398) = 14.06, $p < .001$, accounting for 3.4% of the variability in score changes with adjusted R2 = 3.2%. This is a weak predictive relationship (Cohen, 1988). The correlation between amount of listening time and score improvement was statistically significant, $r(398) = .185$, $p < .001$. This result needs to be viewed suspiciously as the data were considerably skewed due to the large number of participants who did not listen while they were engaged in ER.

Q4: To what extent did the attained reading speed affect changes in TOEFL ITP scores?

A linear regression was conducted to examine how well the average reading speed could predict improvements in TOEFL ITP scores. A scatter plot showed that the relationship between reading speed and score improvement did not reveal any bivariate outliers. An analysis of standard residuals showed that the data contained no outliers (Std. Residual Min. = -3.02, Std. Residual Max. = 2.95). Independence of residual errors was confirmed with a Durbin-Watson test ($d = 1.875$). Residual plots showed homoscedasticity and normality of the residuals. The reading speed did not significantly predict score increases, F (1, 398) = 0.185, $p = .667$, accounting for 0% of the variability in score changes with adjusted R2 = 0.02%. This is a negligible predictive relationship (Cohen, 1988). The correlation between reading speed and score improvement was not statistically significant, $r(398) = –0.002$, $p = .667$.

Q5: To what extent did the average graded reader book level read by learners affect changes in TOEFL ITP scores?

A linear regression was conducted to examine how well the average graded reader book level could predict improvements in TOEFL ITP scores. A scatter plot showed that the relationship between book level and score improvement did not reveal any bivariate outliers. An analysis of standard residuals showed that the data contained no outliers (Std. Residual Min. = -2.99, Std. Residual Max. = 2.95). Independence of residual errors was confirmed with a Durbin-Watson test ($d = 1.773$). Residual plots showed homoscedasticity and normality of the residuals. The book level did not significantly predict score increases,

F (1, 398) = 0.101, p = .751, accounting for 0% of the variability in score changes with adjusted R2 = 0.02%. This is a negligible predictive relationship (Cohen, 1988). The correlation between the book level and score improvement was not statistically significant, $r(398) = –0.002$, p = .751.

Q6: Did the top readers (by volume) or the top test score improvers show a connection between reading and test scores?

To answer this question two comparisons were made. First, the top 50 readers were compared with the bottom 50 readers in the data set. The top readers, with an average word count of 219,090 and having spent an average of 29.1 hours reading and 10 hours listening, experienced an average TOEFL ITP score increase of 30.8 points. At the same time, the bottom 50 readers, with an average word count of 29,237 and having spent only 4.7 hours reading and 0.7 hours listening, had an average TOEFL ITP score increase of 11.5 points. This would seem to suggest a strong benefit for doing ER. Next, the ER reading performance of those students who showed score increases (n = 281) were compared with students whose scores had actually decreased during the same period (n = 86). The former group had read an average of 115,170 words, slightly higher than the mean for the entire group (107,892 words), while the latter group had read an average of 88,629 words. Lower, but not tremendously so. A trend can be clearly seen here, but it is not robust.

No strong or even moderate correlation or predictors emerged from the analysis. From these results, we can draw the conclusion that larger word counts and more time spent reading (obviously connected factors) were responsible for a small, but significant, amount of variance in TOEFL ITP score changes between April and December. To a lesser extent, listening while reading also seemed to predict score improvement, though the large number of students who either did not engage in this activity or did so only minimally, means that the result should not be taken as proof that listening while reading contributes significantly to test score improvement, though it hints in that direction. Neither reading speed nor average book level showed any statistically significant impact on score improvement.

Discussion

While the results of this study show a significant effect for ER on score increases, the amount of this effect is not especially large. Achieved word counts and time spent reading are correlated with higher scores, and turning on the audio and listening while reading may also be. In this regard, the current study is in line with previous studies on ER and TOEFL ITP score improvement (Van Amelsvoort, 2019, and Williams, 2009) and seems to be approaching agreement on the effect of adding listening to ER (Milliner, 2019).

At the same time, however, it seems that some other issues are raised by the results.

Table 2. Summary of effect sizes

	R	R^2	% of variance	p
Word count	0.223	0.050	5%	$p < .001$
Reading time	0.254	0.065	6.5%	$p < .001$
Listening time*	0.185	0.034	3.4%	$p < .001$
Reading speed	0.022	0	0%	$p = .667$
Avg. book level	0.016	0	0%	$p = .751$

Note. * failed to achieve homoscedasticity.

First of all, it seems that extensive reading is unlikely to result in large TOEFL ITP gains at scale in a period of one academic year given the language and content of graded readers. Van Amelsvoort (2019) explained that the TOEFL ITP is written at a CEFR level of roughly C1. For ER, most students in the current study were reading at an A1-A2 level, though a few occasionally read at a B1 level. There is still a considerable gap between the level at which the students were able to read fluently, and that of the test, and not enough time is available to significantly close it. This gap consists of complex grammar constructions and a large amount of infrequent vocabulary, neither of which is found in the books students are reading. Reading rate improvements with ER may only provide indirect help with some parts of the test, for example, the listening section, where speeds of 120-215 words per minute are needed to be able to read the questions and possible answers in the time allotted, but test success is still dependent upon accurately decoding and recognizing aural input. Likewise, the structure section and reading section rely considerably on grammar knowledge, discourse knowledge, background knowledge, etc., in addition to reading fluency.

Second, longitudinal evidence from the current program suggests that the content of reading lessons and the specific vocabulary students are encountering may have a greater effect on their performance on the test than reading fluency. Despite the existence of an ER program for years and relatively consistent TOEFL ITP scores, after the current program switched to a liberal arts program with intensive reading and semi-integrated ER, score gains increased (Wadden, Onoda, & Van Amelsvoort, 2023). For this new syllabus, ER was adapted and given a slightly different role. Instead of all books being chosen by students, weekly assigned books were required that were thematically connected to the reading content of the modules. This allowed students to repeatedly encounter words and concepts which they would see in regular lessons.

The results in this study may underline the limitations of stand-alone ER and the need to combine it with other reading activities to increase effectiveness. An effective TOEFL ITP program will likely feature intensive reading with narrow fluency (timed reading practice and oral repeated reading) along with ER for both narrow and wide fluency training (Milliner, 2021, Shimono, 2023). To these, test familiarization and strategy instruction would also likely be helpful if added.

Limitations

As with other ER studies, difficulties in controlling student exposure to language beyond ER apply, though hopefully to a lesser extent. Student cheating also likely corrupted the data, and a larger total volume of reading would have made the study's conclusions stronger. The study did not distinguish between less and more proficient students, which also may have shown that important differences exist between these groups.

References

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language, 26,* 29-49.

Carney, N. (2016). Gauging extensive reading's relationship with TOEIC reading score growth. *Journal of Extensive Reading 4*(4). https://jalt-publications.org/content/index.php/jer/article/view/11/9

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

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.

ETS Japan (n.d.). TOEFL ITP usage in Japan. Retrieved from https://www.toefl-ibt.jp/educators/search/toefl_itp/

Extensive Reading Foundation (n.d.). ERF graded reader scale. Retrieved from https://erfoundation.org/wordpress/graded-readers/erf-graded-reader-scale/

Gobel, P. (2011). The effect of reading while listening on TOEFL gains. *Koto Kyoiku Forum, 1*, 45-51

Gradman, H., & Hanania, E. (1991). Language learning background factors and ESL proficiency. *Modern Language Journal, 75*(1), 39-51.

Lyddon, P. A., & Kramer, B. (2019). Connecting extensive reading to TOEIC performance. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouësny (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 257- 262). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.38.1019

Milliner, B. (2019). Comparing extensive reading to extensive reading-while-listening on smartphones: Impacts on listening and reading performance for beginning students. *The Reading Matrix, 19*(1), 1-19. https://www.readingmatrix.com/files/20-81br6g10.pdf

Milliner, B. (2021). The effects of combining timed reading, repeated oral reading, and extensive reading. *Reading in a Foreign Language, 33*(2), 191-211.

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

Nation, I.S.P., (1997). The language learning benefits of extensive reading, *The Language Teacher, 21*(5), 13-16

Nishizawa, H., Yoshioka, T., & Fukada, M. (2010). The impact of a 4-year extensive reading program. In A. M. Stoke (Ed.), *JALT2009 conference proceedings* (pp. 632-40). JALT. https://jalt-publications.org/archive/proceedings/2009/E035.pdf

Nishizawa, H. & Yoshioka, T. (2016). Longitudinal case study of a 7-year long ER program. In M. Gobert (Ed.), *Proceedings of the 3rd world congress on extensive reading* (pp. 28-40). Leanpub.

O'Neill, B. (2012). Investigating the effects of Extensive Reading on TOEIC reading section scores. *Extensive Reading World Congress Proceedings, 1*, 30-33. https://erfoundation.org/proceedings/erwc1-ONeill.pdf

Robb, T., & Kano, M. (2013). Effective extensive reading outside the classroom: A large-scale experiment. *Reading in a Foreign Language, 25*(2), 234–247.

Rutson-Griffiths A., & Rutson-Griffiths, Y. (2018). The relationship between extensive reading and TOEIC score gains. *Hiroshima Bunkyo Women's University Journal of Higher Education, 4*, 41-50.

Shimono, T. R. (2023). The effects of extensive reading, timed reading, and repeated oral reading on Japanese university L2 English learners' reading rates and comprehension over one academic year. *Reading in a Foreign Language, 35*(2), 190-221. https://hdl.handle.net/10125/67447

Storey, C., Gibson, K., & Williamson, R. (2006). Can extensive reading boost TOEIC scores? In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.) *JALT2005 conference proceedings* (pp. 1004-1018). JALT. http://jalt-publications.org/archive/proceedings/2005/E034.pdf

Van Amelsvoort, M. (Aug. 12, 2019). *How helpful is ER fluency training for improving TOEFL ITP scores?* Presentation at 5th World Congress on Extensive Reading, Taichung, Taiwan.