*The Reading Matrix: An International Online Journal*
*Volume 19, Number 1, April 2019*

# Comparing Extensive Reading to Extensive Reading-While-Listening on Smartphones: Impacts on Listening and Reading Performance for Beginning Students

**Brett Milliner**
Tamagawa University

## ABSTRACT

*This study traced the development of beginner-level (CEFR A1 and A2) Japanese university English learners' listening and reading skills (N=58). Improvements in listening and reading performance were compared among three groups of low proficiency EFL students: (1) learners who simultaneously read while listening to over 100,000 words from graded reader audiobooks (n=19), (2) learners who read over 100,000 words from digital graded readers (n=17), and (3) a control group who were not required to do any extensive reading (n=22). Changes in listening proficiency were evaluated by: (1) changes in TOEIC® listening test scores, and, (2), changes in listening vocabulary levels test (LVLT) results, while developments in reading proficiency were evaluated by: (1) changes in TOEIC® reading test scores. While the reading-while-listening group achieved significant gains across all measurements, and in comparison to the reading group, achieved better post-treatment results, the strong performance from the control group makes it difficult to conclude that extensive reading-while-listening is the most effectual approach for lower proficiency EFL learners.*

## INTRODUCTION

In foreign language learning environments, learners may often benefit from increased language input. A popular vehicle for overcoming this input-deficit is extensive reading (ER), which has also been shown to promote reading fluency, language acquisition and a range of other language skills. However, would an additional mode of input be helpful? That is to say, what would happen if learners listened while reading a text? Could the addition of listening input help promote gains in listening and reading fluency?

In addition to ER, researchers have advocated extensive watching (e.g., Saunders & Ishimaki, 2016), extensive listening (Milliner, 2017; Renandya & Farrell, 2011; Stephens, 2010) and simultaneous reading-while-listening (defined as extensive listening by Chang & Millett, 2014 & 2016) as practices which increase target language input. In comparison to ER, however, these approaches have received limited attention from researchers and a dearth of empirical data in this field is reported (Chang & Millett, 2016). This paper investigates whether beginner-level English language learners (A1 and A2) simultaneously reading-while-listening (RWL) to graded reader audiobooks can experience improvement in listening and reading proficiency as reflected in

TOEIC test scores, and changes in LVLT results, and whether such improvements are greater than any experienced during graded reading without audio input. More concisely, this study tests the hypothesis that extensive RWL is a more effective approach for developing beginner-level students' receptive skills.

**Extensive Reading**

Most ER programs share a common purpose: that learners read large quantities of self-selected, simplified texts in an environment that promotes the enjoyment of reading in a foreign language (Day & Bamford, 1998; Renandya, 2007). Reflecting the growing body of research advocating the benefits of ER, and a general recognition from foreign language teachers that their students need to encounter more foreign language input, extensive reading components can now be found in many foreign language programs. Researchers conducting empirical studies such as Beglar, Hunt, and Kite (2012) claim to have demonstrated ER's efficacy over other approaches - such as intensive reading - for developing second language students' reading skills. In a meta-analysis of 34 research studies in ER, Nakanishi (2015) concluded that ER "improves students' reading proficiency and should be a part of language learning curricula" (p.6). In another contemporary analysis of ER research Jeon and Day (2015) considered treatments involving five core components: *easy reading material*; *learners choose what they want to read*; *learners read as much as possible*; *reading is individual and silent*; and, *teachers orient and guide their students* (p. 302). Their review concluded that reading programs with this focus were able to achieve greater improvements in reading proficiency than intensive or traditional reading approaches.

At this juncture, however, it is important to note that the reliability of ER research has been questioned recently. Key concerns include: (1) how the ER treatment is being conceptualized by the teacher/researcher (Waring & McLean, 2015), (2) the lack of long-term treatments (Nakanishi, 2015), (3) a lack of clarity on how much reading was actually done (Waring & McLean, 2015). It is therefore important, moving forward, that these points are addressed if one is to objectively measure the effects of ER and the listening (RWL) treatments now to be considered. How this study addresses these concerns will be described in the research methodology section later.

**Reading-While-Listening (RWL)**

As was noted in the introduction, within ER research, there has been limited interest in the impact simultaneously reading-while-listening (RWL) can have upon the learner's reading and listening skills. While this paper addresses the latter, from the research completed thus far, there are some encouraging signs that RWL can foster both reading and listening fluency and that it may be particularly beneficial for beginner level English learners.

Focusing on Japanese university EFL learners, Gobel (2011) reported on 162 freshman students (CEFR levels A1-B2) reading-while-listening (RWL) to graded reader audiobooks for one year. Comparing students' TOEFL scores from before and after the treatment, Gobel found that the more students practiced RWL, the higher their TOEFL test gains were. Gobel also considered whether the level of material students used influenced TOEFL gains. This analysis revealed the level of material had only a weak correlation with TOEFL gains and he argued that the reading of higher level texts did not lead to significant vocabulary gains, but did lead the students to processing denser texts more efficiently. Gobel (p. 50) did not include an exact figure for how much RWL was completed, but instead provided "a rough calculation that students read

for a total of eight hours over the year-long treatment (17,400 words)." While Gobel noted that this volume of RWL could hardly be called extensive, he highlighted the impressive gains some students made after minimal amounts of RWL.

Two researchers who have investigated a range of factors related to extensive listening and RWL are Anna Chang and Sonia Millett. In 2015, the authors compared reading fluency gains between RWL (which they called audio-assisted reading) and silent reading treatment groups. Over 26-weeks, a sample of 64 beginner-level Taiwanese EFL students either read or practiced RWL using 10 level one and 10 level two graded readers from the Oxford Bookworms series (117,401 words in total). Judging by results from a pre, post and delayed reading test based on other graded reader titles (from the Oxford Bookworms catalogue), the RWL group achieved a much larger improvement in reading speed after the post and delayed-post-tests. The RWL students also achieved significantly higher gains in reading speed between the pre-treatment and delayed-post-treatment measurements. The authors posited that increased reading speeds were due to the audio input acting as a reading speed pacer, and helping readers stay on task. Moreover, the researchers wondered whether learners were being drawn more deeply into stories through sound effects and engaging narration found in the audiobooks. After testing for development in reading comprehension levels, results indicated that both treatment groups achieved increases, and their comprehension skills had not deteriorated by the time of the delayed post-test. However, in a comparison between treatment groups, the RWL group made a substantially greater improvement. In their discussion, Chang and Millett compared the effects of RWL with other reading rate enhancement activities to conclude that RWL has a stronger impact on enhancing reading comprehension than timed reading, repeated reading, or silent extensive reading.

In 2016, Chang and Millett focused on listening skill development for 76 beginners to low-intermediate students studying EFL at a Taiwanese university. The researchers were interested in the development of student's listening fluency after they: (1) used RWL with a graded reader in the classroom, and (2) listened to the same graded reader and answered post-listening comprehension questions for homework. By the end of the treatment, the English learners had read a total of 15 books (at a total word count of 85,712 words). The researchers concluded that the inclusion of post-listening focused activities after ER or EL (extensive listening) can significantly improve English learners' listening competence, leading to higher TOEIC listening scores. Another important finding from this study was that beginning-level listeners took longer to listen effectively, and the researchers recommended they spend more time engaged in RWL to beginning-level texts to become more familiar with listening and reading in English. Lastly, the researchers observed that the more their students practiced, the better they became and that the post-listening activities seemed to promote students' listening competence.

In 2012, Chang compared the effects of RWL (n=31) and intensive listening (n=24) treatments for Taiwanese college students. Listening fluency gains were evaluated using pre- and post-treatment TOEIC listening tests and Vocabulary Levels Test scores. Students in the RWL treatment read 15 graded reader titles (selected by Chang) from the Oxford Bookworm series during class time over the 26-week treatment. Students in the intensive listening group focused on reading and understanding three books from the same Oxford series. After the treatment, only the RWL group made a significant gain in TOEIC listening test scores while the intensive listening group made a significant gain in the Vocabulary Levels Tests. Reflecting on results from another a small-scale study she conducted in 2011, Chang posited that reading-while- listening to 15 easy graded reader titles may not be sufficient to bring about significant low-frequency vocabulary gains.

One concern with a simultaneous RWL approach is that learners' reading speeds or listening competencies may not be at a level to perform this task effectively (Chang, 2009; Chang & Millett, 2015; Chang & Millett, 2016; Gobel, 2011). As was noted by Chang (2009) and Gobel (2011), one of the biggest difficulties L2 listeners face is that the speech rate is too fast. Reading rates among L2 learners are reported to be very slow (Beglar et al., 2012; Chang & Millett, 2015; McLean & Rouault, 2017), so they may not be able to keep up with the audio input. In the case of most of Chang and Millett's treatments, students were not allowed to move through books at their own pace by slowing down the audio input to match their reading speed. Despite these conditions, the authors observed some encouraging results. However, what would happen if students had control over these factors? Would allowing students to adjust the speech rates produce more promising results?

Chang and Millett have made a tremendous contribution to RWL, EL and ER research and they have greatly influenced the design of this study, but their studies can be criticized for not mirroring some of the core and variable guidelines for ER research (Waring & McLean, 2015). As was also noted by the authors (e.g., Chang & Millett, 2015) limitations with their research design included: participants not being allowed to freely choose reading texts, not allowing readers to move through books at their own pace, and experiencing the majority of ER or RWL in a controlled classroom environment. Also, the actual volume of RWL or listening done is questionable. Although Waring and McLean have included students reading in large volumes as a core component of an extensive reading program, they stopped short of specifying how much is enough. If one follows the guidelines advocated by Beglar and Hunt (2014), which specify that EFL students need to read in excess of 200,000 words a year to promote significant reading fluency and vocabulary acquisition, the English learners in Chang and Millet's studies appear to have not reached the input threshold. For example, Chang (2012) noted that her students for listening during a 26-week treatment used 15 titles. Additionally, in Chang and Millett's (2016) study, students read 85,712 words, and according to Chang and Millett (2015), after a 26-week period it was calculated that each student had read an average of 117,401 words.

**The Current Study**

By evaluating an ER and RWL intervention more closely following the guidelines for extensive reading research established by Waring and McLean (2015), this report is attempting to provide a fairer empirical review of ER and RWL for low-level English learners. Moreover, in light of the abovementioned positive influences RWL can have on lower-level English learners' reading and listening fluency (Chang & Millet, 2015), this study aims to provide another evaluation of a RWL approach, this time focusing on beginner-level Japanese university students using digital stories.

In recognition of the concern that L2 learners' slow reading speeds or listening skills may prevent them from performing RWL effectively, an important divergence in this study's design is that learners could control the speed of English input. Students read digital books or audiobooks within the Xreading system (xreading.com). The use of this system (with appropriate teacher facilitation) allowed learners to read at their own pace and adjust the speech rate to meet their listening and/or reading competencies. To receive the maximum score dedicated to ER for the course (10%), students needed to read/RWL to more than 100,000 words in the semester (April 2016 - July 2016).

Focusing on receptive fluency development, and whether reading-while-listening with graded readers would transfer to wider listening and reading skill development, this evaluation set out to answer the following research questions:

1. To what degree did the students' TOEIC listening section scores change between the start and end of the treatment period?
2. To what degree did the students' TOEIC reading scores change between the start and end of the treatment period?
3. To what degree did the students' listening vocabulary levels test (LVLT) scores change between the start and end of the treatment period?

## METHODOLOGY

### Participants

This study was undertaken in the researcher's compulsory four-skills English classes at a private university in Tokyo, Japan. All participants were freshman Japanese students aged between 18 and 21 years old studying in the Education (17), Agriculture (21), or Liberal Arts (22) departments. All students were streamed into their beginner-level, English class (correlating with CEFR level A1) based on TOEIC Bridge or TOEIC IP placement test scores. The average pre-test TOEIC score was 290. Only students who actively participated in all classes and whom reached the 100,000-word reading target were included in the evaluation.

The researcher followed ethical guidelines for research and approval from the university's research committee was received before commencing this study. Students also signed a written consent form, which permitted the researcher to access their test scores and Xreading data for analysis purposes.

### Xreading

For the facilitation of EL and ER, the program Xreading (xreading.com) was adopted. Xreading is an online library of graded readers and learning management system (LMS) devoted specifically to ER. There are over 500 titles in the Xreading library for reading purposes and most titles have an audiobook function, so learners are able to listen or simultaneously read-while-listening to a book. An individual license (¥2,400/$22US annually) grants access via a mobile device or PC to all books in the library, as well as analytics data (e.g., reading speed, total words read, books read, and reading times) and post-reading quizzes.

### Treatment: How ER and RWL were implemented

The treatment period lasted 15 weeks. The researcher taught all classes and they met twice a week for 100 minutes. Textbooks, materials and grading procedures for each of the three groups were the same, except for the extensive reading component. The textbooks were Successful Keys for the TOEIC Test Intro (Mizumoto & Stafford, 2014) and Read This (Mackey, 2010). Approximately half of each textbook was completed during the 15 weeks. For the ER and RWL groups, at the start of each class, ten-minutes were dedicated to silent reading, book discussions, follow-up ER or RWL training. In order to reach the monthly reading (word) targets, students were expected to read outside of class time. To more clearly describe how ER and RWL were

implemented in this study, the researcher adapted Waring and McLean's (2015, p.165) list of core and variable attributes for extensive reading research (see Table 1 below).

**Table. 1** Summary of ER core and variable dimensions for this study (adapted from Waring & McLean, 2015)

| **Core Elements** | **Managed in this study by:** |
|---|---|
| Fluent, sustained comprehension | Analysis of reading speed, post reading quizzes, ER awareness exercises were employed to promote fluent comprehension of texts (Brierley, Ruzicka, Sato & Wakasugi, 2010) |
| Large volume of material | ● Students were asked to read or listen to more than 100,000 words to receive a maximum score dedicated to ER/RWL in the course:<br>● To ensure that students read consistently, smaller monthly word targets were set |
| Reading over extended periods of time | ● Students accessed graded readers/audiobooks via their personal smartphone or personal computer inside or outside classroom<br>● Reading/Listening time was recorded by the Xreading system |
| Longer texts | ● Students chose books at a level appropriate for them |
| **Variable Elements** | **Managed in this study by:** |
| ER/RWL is conducted at home or in class | ● 10 minutes of class time was dedicated to ER/RWL and technical training during a 30-class (twice-weekly) treatment period<br>● Students were instructed to engage in ER or RWL in their own time in order to reach reading targets |

| | |
|---|---|
| ER/RWL is required | Participation in this ER/RWL program was not compulsory. However, students who failed to participate forfeited 10% of their final grade |
| ER/RWL is monitored | ● ER/EL was monitored by the Xreading system<br>● Student's WPM, level of book, post-reading quiz results and progressive word counts were scrutinized throughout the treatment period<br>● Students were not allowed to proceed to their next book unless they passed the post-reading quiz |
| The reading/listening is assessed | Students received 10 points (10%) towards their overall grade if they reached the 100,000 word target |
| Reading follow-up activities | ● Post-reading/listening comprehension quizzes (conducted inside Xreading)<br>● Follow-up training on how to use the website effectively<br>● In-class discussions concerning interesting titles, reading analytics, book levels & ER |
| Graded or non-graded readers used | Graded-readers from the Xreading library (no restrictions were placed on the level of book) |
| Longer or shorter texts | Students were free to choose texts of any length in the Xreading library |
| Freedom to select texts | ● Students were free to choose titles from the Xreading library<br>● No book could be used twice |

For the control group, the first ten minutes of classes were dedicated to a five-minute quick writing and connected speaking fluency activity (adapted from Nation & Newton, 2009). Students were also asked to rewrite their quick writing compositions and a class reflection for homework

(minimum 150 words). Students in the control group covered the same textbook content and all assessment tasks were the same as the treatment groups. Concerning grading, while the ER and RWL students received 10% for their reading efforts, this 10% was shared between student's classwork grade (+5%) and intensive reading assessment (+5%)

## Instruments

Students read or undertook RWL using their personal laptop computer, tablet or smartphone and their reading analytics data was collected by the Xreading system. This data, for example word counts and reading times, was analyzed by the researcher both during and after the treatment period.

In an attempt to gauge student improvement in reading and listening skills, TOEIC test scores for tests taken before (April 2016) and after the treatment (July 2016) were considered. Some students had been streamed into the elementary courses based on their scores in a TOEIC bridge placement test (a shortened form of the TOEIC test that students sat before entering the course in April 2016). For comparison purposes, students' TOEIC Bridge data was converted to TOEIC scores using an official TOEIC® Bridge and TOEIC comparisons chart (ETS, 2016). At the end of the treatment, all students sat a TOEIC test, which was a course requirement with TOEIC scores accounting for 20% of students' final grade.

To determine changes in aural vocabulary growth, students completed the LVLT- listening vocabulary levels test (McLean, Kramer & Beglar, 2015). Under the supervision of the researcher, students undertook the first test during the second class of the course. The researcher did not disclose the answers because the test was used again as a post-treatment test. The post-treatment test was held in the last class of the course, and upon completing the assessment; the researcher shared all results with students.

## Procedure

During the first meeting, the researcher introduced the program and the reasons why students would be doing ER or RWL. Subsequently, students were instructed on how to access Xreading and start reading books. Based on the author's experiences leading students to read digital graded readers using the X-Reading system in previous semesters (see Milliner & Cote, 2015), learner training followed a cyclical process, whereby students were initially instructed on how to use the XR website, then periodically throughout the course, portions of the in-class reading time was dedicated to follow-up technical training and discussions on how to undertake digital ER/RWL more effectively. For example, students learned how to select level-appropriate materials, evaluate reading analytics, and exploit the affordances of digital books (e.g., customizing text display and audio speeds).

## Data Analysis

The independent variable was treatment group- RWL, ER, and control (no extensive input). There were three sets of dependent measures, TOEIC listening and reading pre- and post-treatment scores and LVLT pre- and post-treatment scores. All statistical analysis was undertaken using the JASP 0.9.1.0 (jasp-stats.org) software. A mixed factor ANOVA was used to interpret whether there was a difference between group means to establish whether there was an improvement in the

TOEIC listening section (research question one), TOEIC reading section (research question two), and LVLT result (research question three). For the cases where a significant ANOVA was returned, the effect sizes were calculated using Omega squared ($\omega^2$) as it provides the most reliable effect size measure for smaller sample sizes (Goss-Sampson, 2018).

## RESULTS & DISCUSSION

### TOEIC Listening Section Scores

The evaluation of student's scores in the listening section of the TOEIC (Table 2) revealed that the most substantial gains were achieved by the RWL group (20 points) followed by the ER group (11 points). The Control group achieved the smallest gain of six points. Based on the descriptive data plot (Figure 1 below), largest increases were observed among the RWL students.

**Table. 2** Summary of descriptive statistics for TOEIC Listening

| Post Treatment Score | Treatment | Mean | SD | N |
|---|---|---|---|---|
| TOEIC Listening-1 | Control | 149.5 | 12.14 | 22 |
| | RWL | 204.4 | 15.80 | 17 |
| | Reading | 187.1 | 30.85 | 21 |
| TOEIC Listening-2 | Control | 155.7 | 41.44 | 22 |
| | RWL | 223.5 | 36.18 | 17 |
| | Reading | 202.1 | 37.27 | 21 |

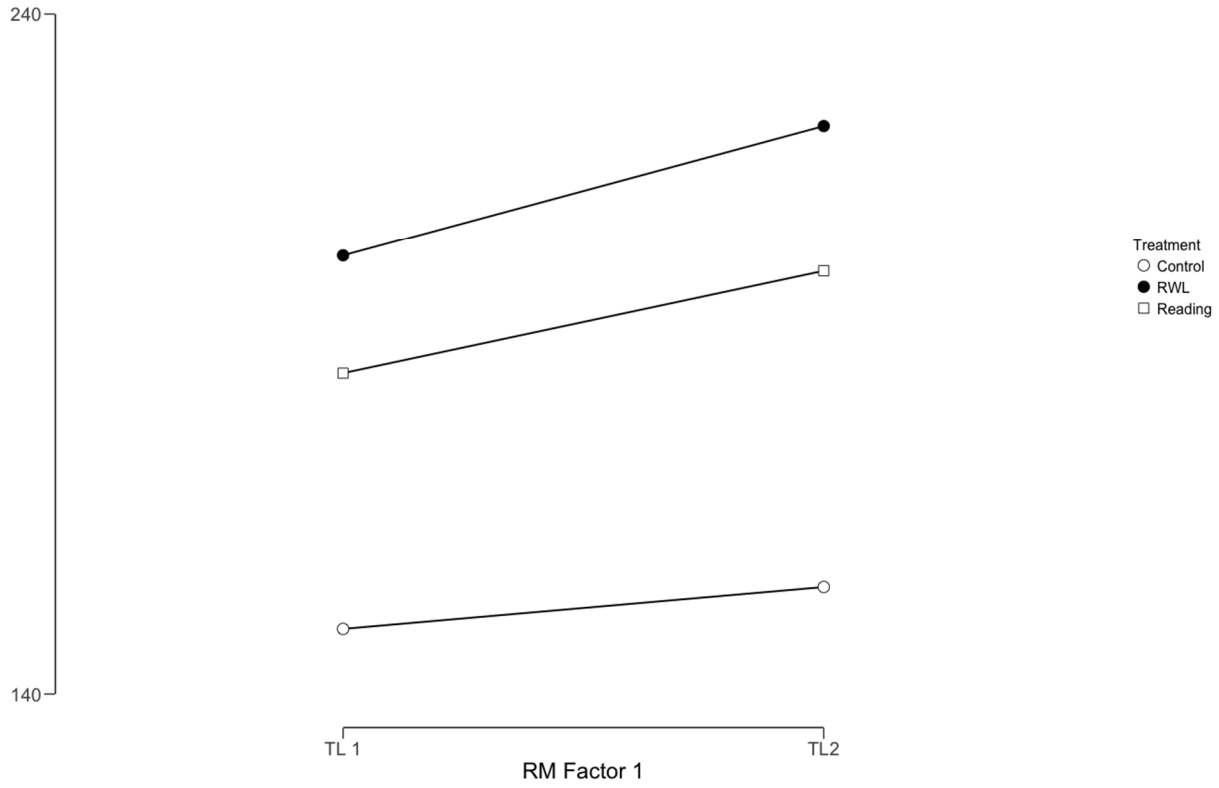


**Figure 1.** Plot of interaction of the three group's TOEIC Listening scores

The mixed factor ANOVA returned a significant main effect for TOEIC listening scores ($F(2,58)= 6.641$, $p<.05$). The Bonferroni corrected post hoc test showed that there was a significant increase in TOEIC listening score irrespective of treatment ($p<.05$). The between subjects factor for treatment type ($F(2,57)= 34.39$, $p<.001$) showed that TOEIC listening scores were affected by the treatment received which also had a large effect size (0.527), however, a significant interaction effect (i.e., between the type of treatment and TOEIC listening scores) was not observed. This makes it hard to confidently conclude that either of the treatments was more effective in promoting better scores in the second TOEIC listening test.

**Table. 3** Summary of within subjects ANOVA

| | **Sum of Squares** | **df** | **Mean Square** | **F** | **p** | **$\omega^2$** |
|---|---|---|---|---|---|---|
| TOEIC Listening Score | 5333.8 | 1 | 5333.8 | 6.641 | 0.013 | 0.039 |

| | | | | | | |
|---|---|---|---|---|---|---|
| TOEIC Listening Score ∗ Treatment | 878.1 | 2 | 439.1 | 0.547 | 0.582 | 0.000 |
| Residual | 45779.2 | 57 | 803.1 | | | |

*Note.* Type III Sum of Squares

The Bonferroni post hoc tests showed each group was able to increase their TOEIC listening score by the end of the treatment ($p<.05$). Following the before mentioned plot (Figure 1)however, one can see that the greatest gains were achieved by the RWL group.

**Table. 5** Summary of Bonferroni post hoc returns

| **Post Hoc Comparisons - Post Treatment Score** | | | | | |
|---|---|---|---|---|---|
| | | **Mean Difference** | **SE** | **t** | **p bonf** |
| TL-1 | TL-2 | -12.92 | 5.134 | -2.516 | 0.015 |

This result supports the hypothesis that RWL was a more efficient approach for developing beginner students' listening skills than ER. This conclusion also lends support to a similar conclusion reached by Chang and Millett (2016) - an extensive RWL program with graded monologic, narrative-type texts seems transferable to the TOEIC listening test, which features a range of textual complexities and styles of listening input (e.g., dialogic speech and formal announcements).

## TOEIC Reading Section Scores

Comparing the TOEIC reading section scores, gains were achieved by the RWL (12 points) and the Control group (22 points), while surprisingly; the ER group recorded a decrease (15 points).

**Table. 6** Summary of descriptive statistics for TOEIC reading section

| **Post Treatment Score** | **Treatment** | **Mean** | **SD** | **N** |
|---|---|---|---|---|
| TOEIC Reading -1 | Control | 79.77 | 14.35 | 22 |
| | RWL | 122.94 | 14.58 | 17 |
| | Reading | 136.90 | 31.84 | 21 |

| | | | | |
|---|---|---|---|---|
| TOEIC Reading-2 | Control | 100.00 | 27.39 | 22 |
| | RWL | 134.71 | 37.85 | 17 |
| | Reading | 121.67 | 22.82 | 21 |

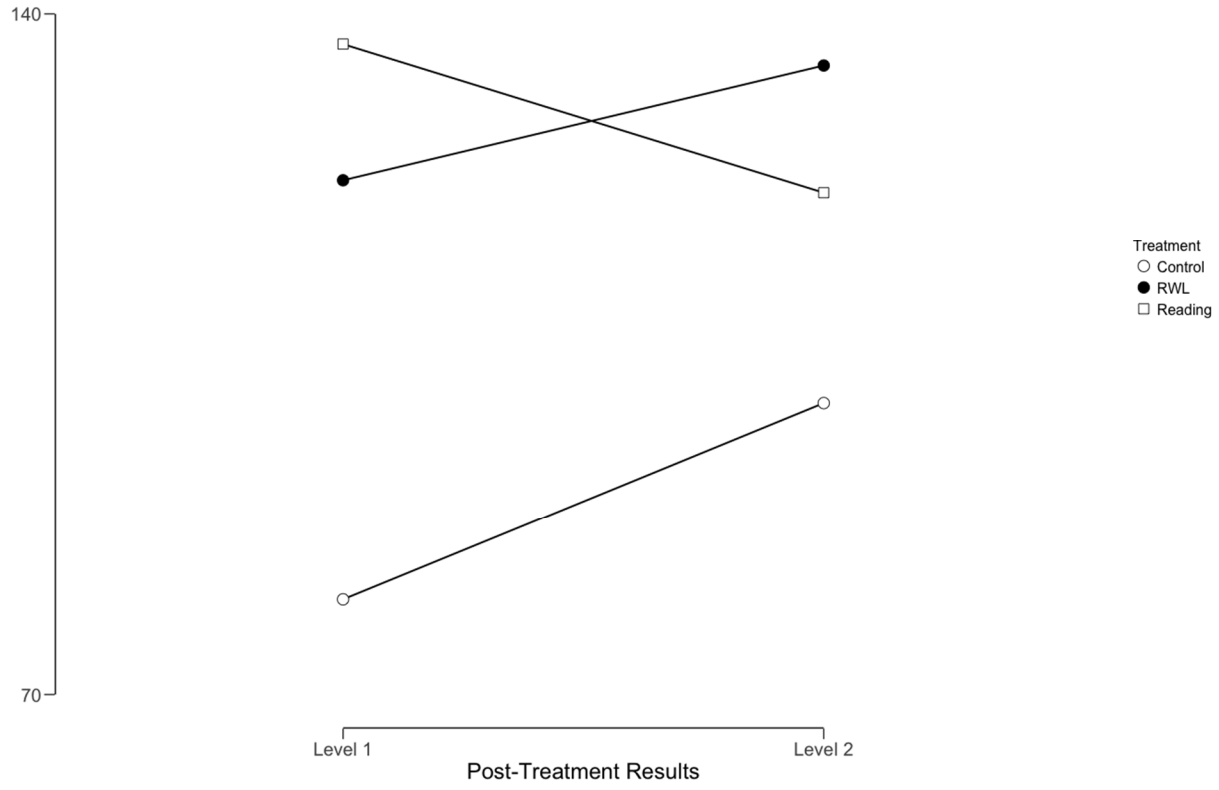


**Figure 2.** Plot of interaction of the three group's TOEIC Reading section scores

The mixed factor ANOVA returned a significant between subjects effect for treatment type ($F(2,57)= 29.56$, $p<.001$), showing that the type of treatment had a strong effect (0.488). Using the Greenhouse-Geisser correction, a significant interaction effect between post-test scores and treatment type ($F(2,57)= 3617.5$, $p=.005$) was also observed, but the effect size was small ($\omega^2=.049$). In the Bonferroni post hoc test, however, a significant increase in reading test score was only observed for the Control group ($p<.001$). For the RWL group then, it is difficult to conclude

that the RWL experience played a significant role in helping students achieve higher TOEIC reading section scores.

**Table. 7** Within subjects ANOVA results for TOEIC Reading section scores

| | Sum of Squares | df | Mean Square | F | p | $\omega^2$ |
|---|---|---|---|---|---|---|
| Post-Treatment Results | 924.0 | 1 | 924.0 | 1.470 | 0.230 | 0.004 |
| Post-Treatment Results ∗ Treatment | 7234.9 | 2 | 3617.5 | 5.756 | 0.005 | 0.049 |
| Residual | 35822.4 | 57 | 628.5 | | | |

*Note.* Type III Sum of Squares

**Table. 8** Between subjects ANOVA results for TOEIC Reading section scores

| | Sum of Squares | df | Mean Square | F | p | $\omega^2$ |
|---|---|---|---|---|---|---|
| Treatment | 42809 | 2 | 21404.4 | 21404.4 | < .001 | 0.488 |
| Residual | 41268 | 57 | 724.0 | 724.0 | | |

*Note.* Type III Sum of Squares

One reason why the ER group failed to achieve a significant gain in reading scores may be related to the volume of reading completed. While reading over 100,000 words in four months is commendable, it is below the 200,000-word benchmark mentioned above for reading fluency development proposed by Beglar and Hunt (2014). And, focusing on improving TOEIC scores specifically, a long-term review of Japanese students engaging in ER over a 120-week period by Nishizawa, Yoshioka and Fukada (2010) proposed learners need to read over 300,000 words to see a marked increase in TOEIC scores. In addition, to the possibility that reading volumes were insufficient, Chang and Millett (2016) noted that students' with low proficiency may take longer than other students to become accustomed to ER or RWL. Both of these factors are particularly relevant to this study which observed beginner-level students reading digital books in their second language for the first time. And, taking these points into consideration, it may be the case that

students needed to read more and become used to ER/RWL in order to achieve a significant increase in TOEIC reading scores.

Nevertheless, it was regrettable to see the ER group achieve an average score 15 points lower post-treatment. This result may be a reflection that the treatment conditions (i.e., where students personally selected graded readers to read at their own pace) did not efficiently promote reading speeds. In Chang and Millett's 2015 study of beginner learners, their RWL group increased reading speed by 47 words per-minute (wpm) while the ER group only improved by an average of 13 wpm. Consequently, the researchers argued that RWL is a superior approach for promoting beginner level students' reading comprehension and reading fluency, so perhaps it is not surprising that only the RWL treatment group was able to improve. It is lamentable that a test of reading fluency, similar to one completed by McLean and Rouault (2017) that tested both reading speed and comprehension was not undertaken in this experiment. In future studies, exploring the efficiency of both approaches (i.e., ER and RWL) at promoting reading fluency would be worthwhile.

## LVLT Scores

The LVLT is a 150-item test designed to measure L2 English learners' aural vocabulary knowledge. Upon completion of the post-treatment round of the LVLT, all groups achieved gains: Control (12.96 points); ER (11.62); and, RWL (12.76). These results are summarized in the plot below (Figure 3).

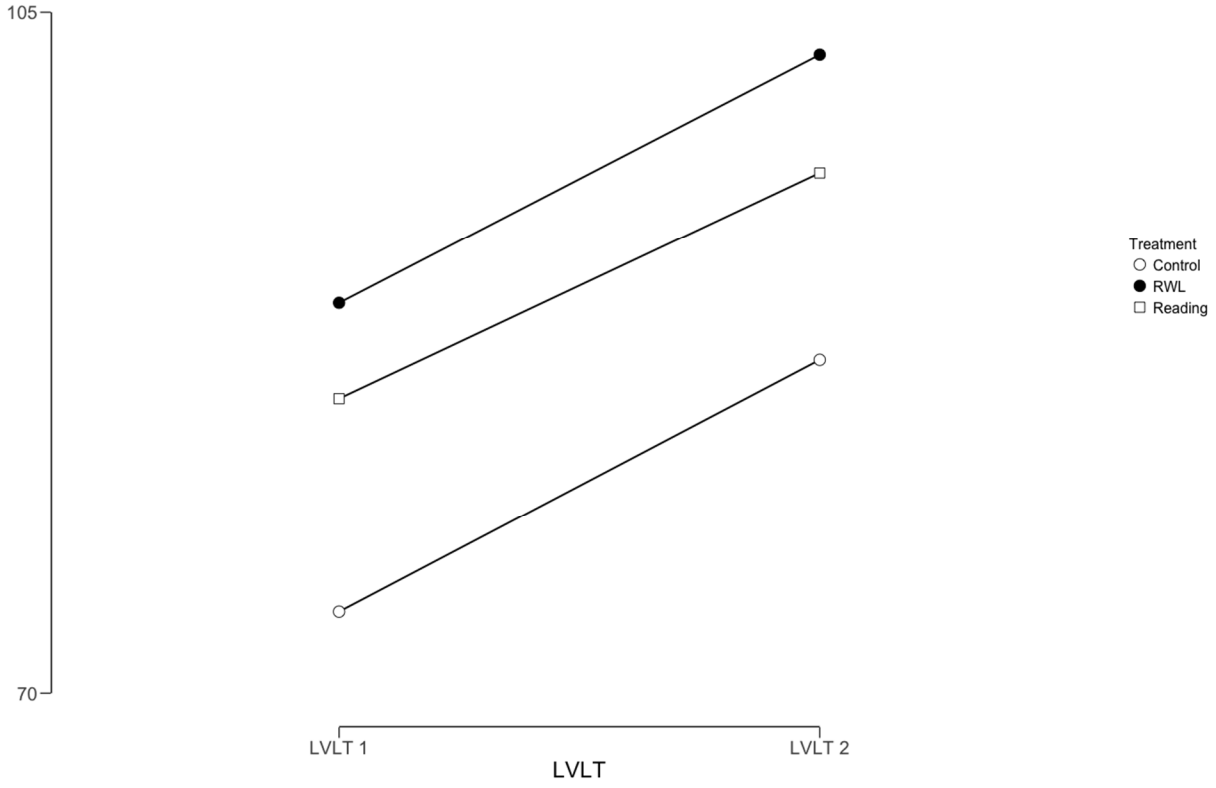

**Figure 3.** Plot detailing the each group's performance between pre- and post-LVLT tests

The mixed factor ANOVA returned a significant main effect for LVLT scores ($F(2,57)= 72.61$, $p<.001$), showing that the type of treatment had a small effect (0.153). Using the Greenhouse-Geisser correction, a significant interaction effect between post-test scores and treatment type ($F(2,57)=6.92$, $p<.01$) was observed, but the effect size was small ($\omega^2=.165$). Following the descriptive data summarized in the plot above, all groups achieved similar increases in LVLT scores with the largest gains observed in the Control group. The Bonferroni post hoc test, revealed that these increases were also significant ($p<.001$).

These results for the LVLT came as a surprise to the researcher, who assumed the RWL or ER groups would be in the best positions to improve in this assessment. A review of where students achieved the largest gains at each of the vocabulary levels (Table 8) failed to reveal any significant trend between each of the groups. The RWL group did, however, make the largest gains at the 3000, 4000 and 5000 levels, and they achieved similar gains to the other groups in the AWL section. These gains could be related to the frequency students met vocabulary within these bands during their RWL practice, however, in a review of their average level of book read throughout the treatment in Xreading, their reading level was similar to the ER group. The improvement achieved by all groups may be explained by students' familiarity with the test format and questions themselves. And, given the strong showing from the control group, it may be a reflection that ER or RWL are not the most efficient methods for acquiring new vocabulary. Chang (2012) observed a similar result, where the RWL group failed to achieve a noteworthy improvement at any of the vocabulary levels. In a Japan-based study by Brown, Waring and Donkaewbua (2008) that focused on incidental vocabulary acquisition after reading, RWL or listening to graded reader stories, the researchers also observed a low acquisition rate for new vocabulary items in each of their treatment groups. In the researchers' words (p.156), it was "quite a disappointing rate of return considering the effort involved."

**Table. 8** Summary of average LVLT results and standard deviations

| **LVLT (#of items)** | **RWL (17)** | **ER (19)** | **Control (22)** |
|---|---|---|---|
| Pre-test | 90.06 | 85.14 | 74.18 |
| Post-test | 102.82 | 96.76 | 87.14 |
| Gain | 12.76 | 11.62 | 12.96 |
| OVERALL (150) | | | |
| Pre-test | 22.3 | 21.10 | 19.64 |
| Post-test | 22.71 | 22.29 | 21.00 |
| *t* | 1.31 | 2.96** | 2.49* |
| 1000 Level (24) | | | |

| | | | |
|---|---|---|---|
| Pre-test<br>Post-test<br>*t*<br>2000 Level (24) | 18.06<br>19.06<br>1.68 | 17.38<br>19.29<br>3.33* | 13.64<br>15.68<br>2.89** |
| Pre-test<br>Post-test<br>*t*<br>3000 Level (24) | 11.59<br>14.88<br>4.67** | 11.38<br>13.10<br>2.22* | 8.45<br>11.55<br>4.01** |
| Pre-test<br>Post-test<br>*t*<br>4000 Level (24) | 12.12<br>14.76<br>4.55** | 11.76<br>14.24<br>3.55** | 9.32<br>12.14<br>3.39** |
| Pre-test<br>Post-test<br>*t*<br>5000 Level (24) | 12.65<br>14.94<br>3.58** | 12.00<br>13.67<br>2.2* | 10.91<br>13.00<br>3.02** |
| Pre-test<br>Post-test<br>*t*<br>AWL Level (30) | 13.29<br>16.47<br>3.48** | 10.71<br>14.90<br>4.18** | 9.50<br>13.77<br>4.09** |

Note: **p<.01
*p<.05

**Limitations**

There are some limitations of this experiment that ought to be highlighted before drawing conclusions. Firstly, TOEIC test scores were used to measure language skill improvement. The TOEIC test is a measure of business English ability, and it is a crude one at best. This test has been criticized for its inability to measure learner achievement over the short term (Childs, 1995), and its ability to measure both listening (Buck, 2001; Chapman & Newfields, 2008) and reading skills (Alderson, 2000; Carney, 2016). Future research ought to consider alternative measurements of constructs such as listening fluency, self-efficacy, reading fluency and listening comprehension.

A second concern is that participants did most of their ER and RWL outside of class time, and not in a controlled research environment. While some degree of monitoring took place in Xreading (e.g., reading speeds and post-reading quiz results), it is difficult to confirm whether students were engaged in ER or RWL exclusively. During the experiment, there was no way to observe whether students shifted between reading, listening or RWL each time they used the Xreading system. It must also be acknowledged that learning experiences other than those involving ER/ RWL may have had significant impact on post-test scores. The TOEIC test result for example, contributed 20% to students overall grade and higher TOEIC scores are perceived by

Japanese society to be valuable for job hunting, so it is not unlikely that some students studied for this test independently.

In the case of the RWL treatment, only audiobooks were used. Following Renandya and Farrell's (2011) and Stephen's (2010) calls for students to be exposed to different types of aural input, perhaps more significant listening gains could be observed if students listened to a wider range of texts (e.g., dialogic texts or texts featuring non-native accents). Although it appeared that the RWL group was able to transfer their listening skills to the different types of texts presented in the TOEIC test, future research ought to consider students listening to a broader variety of listening texts.

Lastly, and most importantly, the small sample sizes of each group ensure similar results may not be observed in other contexts. This evaluation needs to be undertaken with a much larger group of L2 learners to evaluate the effectiveness of RWL. After learning from this initial study, the researcher is looking forward to refining this research methodology (e.g., including a reading fluency measurement) to conduct a similar review with a much larger number of participants.

## CONCLUSION

This study set out to evaluate whether reading-while-listening (RWL) using graded reader audiobooks was more likely than extensive reading (ER) to positively affect listening performance (as measured by TOEIC listening test scores, and LVLT test scores) and reading performance (as measured by TOEIC reading test scores). The results indicate that reading-while-listening to graded reader audiobooks seems to be more efficacious than silent extensive reading in improving the learner's receptive skills in this study. The small improvements in aural vocabulary levels across all three treatment groups suggest that neither ER nor RWL are very efficacious at fostering growth in this area. Learning from this study, the researcher is looking forward to repeating this review with a larger number of participants and a refined research methodology.

***Brett Milliner*** *is an Associate Professor in the Center for English as a Lingua Franca (CELF), at Tamagawa University. Brett's research interests include computer-assisted language learning (CALL), L2 listening and extensive reading. Brett is the co-editor of The Center for ELF Journal.*

Email: **milliner@lit.tamagawa.ac.jp**

## REFERENCES

Alderson, C. (2000). *Assessing Reading*. Cambridge, UK. Cambridge University Press.

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language, 26*(1), 29-48.

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. doi:10.1111/j.1467-9922.2011.00651.x

Brierley, M., Ruzicka, D., Sato, H., & Wakasugi, T. (2010). The measurement problem in Extensive Reading: Students' attitudes. In A. M. Stoke (Ed.), *JALT 2009 Conference Proceedings,* (pp. 641-650). Tokyo: JALT.

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.

Buck, G. (2001). *Assessing Listening*. Cambridge, UK: Cambridge University Press.

Carney, N. (2016). Gauging Extensive Reading's Relationship with TOEIC Reading Score Growth. *Journal of Extensive Reading 4*(4), 69-86. Retrieved from http://jalt-publications.org/content/index.php/jer/article/download/11/9/

Chang, C-S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System, 37*(4), 652-663. https://doi.org/10.1016/j.system.2009.09.009

Chang, C-S. (2011). The effect of reading while listening to audiobooks*: Listening fluency and vocabulary gain. Asian Journal of English Language Teaching, 19*, 43-64.

Chang, C-S. (2012). Gains to L2 learners from extensive listening: Listening development. Vocabulary acquisition and perceptions of the intervention. *Hong Kong Journal of Applied Linguistics, 14*(1), 25-47.

Chang, C-S., & Millett, S. (2013). The effect of extensive listening fluency: some hard evidence. *ELT Journal 68*(1), 31-40. doi: 10.1093/elt/cct052

Chang, C-S., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System, 52*, 91-102. https://doi.org/10.1016/j.system.2015.05.003

Chang, C-S., & Millett, S. (2016). Developing L2 listening fluency through extended listening-focused activities in an extensive listening programme. *RELC Journal, 47*(3), 349-362. doi:10.1177/0033688216631175

Chapman, M., & Newfields, T. (2008). The 'new' TOEIC. *Shiken: JALT Testing & Evaluation SIG Newsletter, 12*(2), 32-37. Retrieved from http://jalt.org/test/cha_new.htm

Childs, M. (1995) Good and bad uses of TOEIC by Japanese companies. In J.D. Brown & S. O. Yamashita (Eds.). *Language Testing in Japan,* (pp. 66-75). Tokyo,Japan: JALT.

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

ETS. (2006). *TOEIC® Bridge and TOEIC score comparisons*. Retrieved from https://www.ets.org/Media/Tests/TOEIC/pdf/ToeicBridge_Cmprsn.pdf

Gobel, P. (2011). The effect of reading while listening on TOEFL gains. *Forum of Higher Education Research, 1*(1), 44-51.

Goss-Sampson, M. A. (2018). *Statistical analysis in JASP: A guide for students* (2nd ed.). Available from: https://static.jasp-stats.org/Statistical%20Analysis%20in%20JASP%20-%20A%20Students%20Guide%20v1.0.pdf

Jeon, E., & Day, R. (2015). The effectiveness of core ER principles. *Reading in a Foreign Language, 27*(2), 302-307.

Mackey, D. (2010). *Read this! Fascinating stories from the content areas 1*. New York: Cambridge University Press.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research, 19*(6), 741-760. https://doi.org/10.1177/1362168814567889

McLean, S., & Rouault, G. (2017). The effectiveness and efficiency of extensive reading at developing reading rates. *System, 70*, 92-106. https://doi.org/10.1016/j.system.2017.09.003

Milliner, B., & Cote, T. (2015). Mobile-based Extensive Reading: An Investigation into Reluctant Readers, *International Journal of Computer-Assisted Language Learning and Teaching*, *5*(4), 1-15. doi: 10.4018/IJCALLT.2015100101

Milliner, B. (2017). Using online extensive listening to promote language input in an EFL setting. In J. Colpaert, A. Aerts, R. Kern & M. Kaiser (Eds.), *CALL in Context Proceedings* (pp. 575-581). Berkeley: University of California.

Mizumoto, A., & Stafford, M. (2014). *Successful keys to the TOEIC Test intro*. Tokyo: Kirihara Shoten.

Nation, P., & Newton, J. (2009). Teaching ESL/EFL listening and speaking. New York: Routledge.

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

Nishizawa, H., Yoshioka, T., & Fukada, M. (2010). The impact of a 4-year extensive reading program. In A. M. Stoke (Ed.) *JALT 2009 Conference Proceedings*, (pp.632-640). Tokyo: JALT.

Renandya, W. A. (2007). The power of extensive reading. *RELC Journal, 38*(2), 133-149. https://doi.org/10.1177/0033688207079578

Renandya, W. A., & Farrell, T. S. C. (2011). 'Teacher the tape is too fast!' Extensive listening in ELT. *ELT Journal, 65*(1), 52-59. doi:10.1093/elt/ccq015

Romeo, K., & Hubbard, P. (2010). Pervasive CALL learner training for improving listening proficiency. In M. Levy, F. Blin, C. B. Siskin, & O. Takeuchi (Eds.). *WorldCALL: International perspectives on computer-assisted language learning,* (pp. 215-229). New York: Routledge.

Saunders, T., & Ishimaki, K. (2016). Learner attitudes of L2-engaged extensive watching versus extensive reading. *The Center for ELF Journal, 2*(2), 1-14.

Stephens, M. (2010). The primacy of extensive listening. *ELT Journal, 65*(3), 311-313. https://doi.org/10.1093/elt/ccq042

Waring, R., & McLean, S. (2015). Exploration of core and variable dimensions of extensive reading research and pedagogy. *Reading in a Foreign Language, 27*(1), 160-167.