L2 reading rate and word length:
The necessity of character-based measurement

Brandon Kramer
Osaka Jogakuin University & Junior College
Temple University, Japan Campus (Ph.D. Student)
Japan

Stuart McLean
Osaka Jogakuin University & Junior College
Kansai University (Ph.D. Student)
Japan

Abstract

Reading rate, usually measured in words per minute, is a common operationalization of reading fluency in second language (L2) research and pedagogy. However, the impact of word length is often not addressed. This paper presents two studies showing how the number of characters in a text influences L2 reading time, independent of word counts, within classroom-based activities for Japanese university English as a Foreign Language students. In Study 1, students ($N = 160$) read two sets of graded texts manipulated to differ only in the total number of characters. The texts with more characters required significantly more time to read, with a small effect size. In Study 2, the average reading times for students ($N = 27$) throughout a semester-long timed reading course were strongly associated with text length as measured in characters, controlling for differences in word counts. Together these studies support the inclusion of character-based counting units when measuring L2 reading rate or reading amount.

Keywords: reading fluency, Latin square, timed reading, extensive reading, word length

The internal validity of a study is the degree to which a researcher can make inferences based on the results of analyzed data. While dependent on a sufficiently robust research design, internal validity is strengthened by high levels of accuracy and precision within the instruments used for measurement. Accuracy can be considered the degree with which an instrument is measuring what it intends to measure. Precision, on the other hand, refers to the degree of consistency and reliability with which the instrument can capture this target construct. One source of a lack of precision is measurement error, which weakens the results of any statistical analyses, and therefore limits the validity of inferences derived from these results. While some error is unavoidable, any unnecessary lack of measurement precision should be addressed to the greatest degree possible. In this paper, we argue for increased accuracy and precision in reading fluency

measurement within second language (L2) reading research by briefly outlining current conceptualizations of reading fluency, reviewing the operationalization of reading fluency within L2 reading research, and then presenting and discussing the results of two studies designed to investigate how the number of characters within a text, rather than simple word counts, can influence reading fluency measurements with data from Japanese university students studying English as a Foreign Language (EFL).

Literature Review

Defining Reading Fluency

Grabe (2009) defined reading fluency as "the ability to read rapidly with ease and accuracy…and text comprehension is the expected outcome" (p. 291). This requires the automatization of several lower-level (bottom-up) processes so that the working memory is free enough to engage in the higher-level (top-down) processes. Some of the more important subskills of reading classified as lower-level processes are orthographic processing, whereby symbols are recognized and interpreted; phonological processing, whereby the orthography triggers the recall of a particular associated phonology; lexical access, where the best-fit word is chosen based on the highest level of activation; and semantic processing, where the associated meanings of the accessed words are activated. Specifically, lexical access, including the identification of the orthographic and phonological subcomponents, needs to become automatic and efficient before L2 readers can effectively engage in text comprehension via syntactic parsing (Grabe, 2009; LaBerge & Samuels, 1974). Although higher-level reading processes such as strategy choice, inferencing, and comprehension checking can also play an important role for more advanced L2 learners, to read fluently the lower-level processes must become automatized, meaning that the processing occurs without effort, intention, or awareness of it even taking place (Logan, 1997; Segalowitz, 2003). In this way reading fluency in L2 research can be defined as the sufficient automatization of lower-level reading processes to allow for adequate text comprehension.

Operationalizing Reading Fluency

In attempting to operationalize reading fluency and the automatization of required subskills, applied linguists and psycholinguists have approached the construct using sophisticated technology incorporating eye-tracking hardware (for a comprehensive overview, see Conklin, Pellicer-Sánchez, & Carrol, 2018; Winke, Godfroid, & Gass, 2013) and reaction times (for a comprehensive overview, see Jiang, 2012; Marsden, Thompson, & Plonsky, 2018). Many classroom researchers do not have access to these tools when conducting their research. Thus, a more accessible operationalization of reading fluency is necessary. Although defined within the literature as "multiple, interactive cognitive processes" (Beglar, Hunt, & Kite, 2012, p. 666; also see Grabe & Stoller, 2011, p. 11) the construct of reading fluency is often reduced within classroom L2 reading research to simply the amount of text read within a given amount of time.

Reading fluency is not only the speed at which one is able to pass their eyes over the words within a text; it also requires adequate comprehension of that text. Although difficult to pin down precisely, assumptions of adequate text comprehension by L2 readers are typically made using

one or a combination of two approaches. One method involves constructing reading materials using a restricted range of vocabulary that are likely to be known by the reader (Nation, 2013). In this process the readers' knowledge of the most frequent words within large representative corpora are tested, such as by using a Vocabulary Levels Test (e.g., McLean & Kramer, 2015, 2016; Nation, 1990; Schmitt, Schmitt, & Clapham, 2001). Based on the results of the vocabulary test they took, texts are then created or adapted from words within the vocabulary levels of which students have demonstrated mastery. While not in itself sufficient for comprehension, knowledge of 98% or more of the words within a given text gives researchers a rough minimum criterion which allows them to assume that the reader knows a sufficient proportion of the vocabulary necessary to be able to read and comprehend the target text (Hu & Nation, 2000; Laufer, 1989; Laufer & Ravenhorst-Kalovski, 2010).

The second commonly used approach for measuring reading comprehension is the administration of a post-hoc comprehension check following the reading of the text. This comprehension check consists of questions testing global or local understanding of the passage with a criterion score set for sufficient passage comprehension (e.g., Beglar et al., 2012; McLean & Rouault, 2017; Quinn, Nation, & Millet, 2007). Using the previous criteria, adequate reading fluency for timed reading passages and extensive reading is therefore operationally measured as the amount of text read relative to a fixed length of time, usually words per minute (wpm), but only for those students for whom adequate comprehension has been demonstrated (McLean, 2014; Nation, 2013).

Reading Fluency Measurement in Practice

For teachers who recognize the importance of building their students' L2 reading fluency, one treatment often shown to encourage gains, operationalized as increases in reading rate without loss of comprehension (e.g., Chang & Millett, 2013; Shimono, 2018; Tran & Nation, 2014), is the progression through a timed reading course (e.g., Millet, 2005; Nation & Malarcher, 2007; Quinn et al., 2007). These courses are usually comprised of short graded texts followed by comprehension questions. The courses are designed such that all texts contain only highly frequent vocabulary and have roughly the same number of words within them, allowing students to focus only on reading more quickly with each subsequent administration without being slowed down by unknown words.

Another treatment which has shown promise in encouraging gains in reading fluency is extensive reading, here defined as "an approach to the teaching and learning of reading in which learners read large amounts of material that are within their linguistic competence" (Grabe & Stoller, 2011, p. 286). Although surmised to benefit other constructs such as lexical richness or vocabulary depth, extensive reading treatments have been most consistently linked to gains in reading fluency similar to the previously described timed reading courses (e.g., Beglar & Hunt, 2014; Beglar et al., 2012; Huffman, 2014; McLean & Rouault, 2017). In such extensive reading research, the independent variable is quantified by measuring the amount of text read and then verifying the comprehension of that text, either using reports written up for each book or more recently using comprehension questions created for each book and administered through an online system such as M-Reader (mreader.org) or Xreading (xreading.com).

Motivations for the Current Studies

Whether researching timed reading or extensive reading, the amount of text must be precisely measured in order to accurately identify fluency gains or, for extensive reading, to account for time-on-task. Previous extensive reading research has relied on such rough measurements as books (e.g., Hitosugi & Day, 2004; Lake, 2014) and pages (e.g., Yamashita, 2004) as measures of reading amount. Most researchers studying timed reading or extensive reading now use the more precise unit of words in studies investigating reading fluency or reading amount (e.g., Horiba & Fukaya, 2015; Sakurai, 2015; Tabata-Sandom, 2017; Tran, 2012; Tran & Nation, 2014).

How precise is the word as a unit of text amount, however? For example, even in books which have been graded using limited vocabulary and grammar, the amount of text for two sentences with the same number of words can be quite different. In Figure 1, we can see two 10-word sentences, both from graded readers commonly included in extensive reading libraries. While both sentences contain 10 words, the second sentence contains 25% more text if measured in characters (Kramer & McLean, 2013). This is because despite containing the same number of words, the second sentence has a longer average word length as measured in characters per word (cpw).
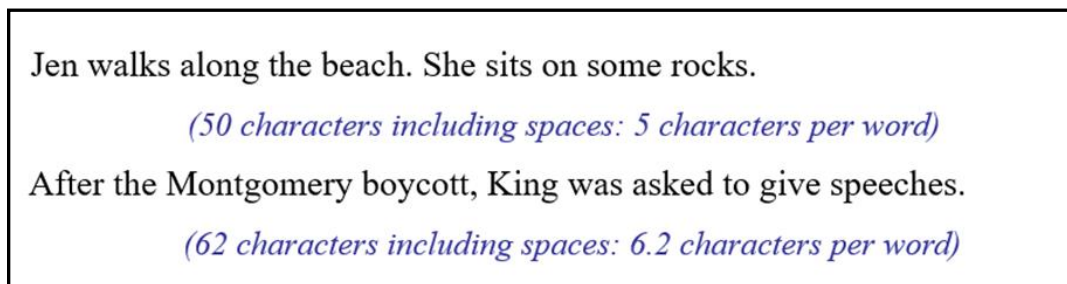
Jen walks along the beach. She sits on some rocks.

(50 characters including spaces: 5 characters per word)

After the Montgomery boycott, King was asked to give speeches.

(62 characters including spaces: 6.2 characters per word)

Figure 1. The difference in the number of characters within 10 words from sentences taken from two books from the same extensive reading library at a university in Japan (adapted from Kramer & McLean, 2013).

The reason for the difference in average word length is clear. The shorter sentence is from an easier book, within MacMillan's Starter Level (Axten, 2008), which contains vocabulary made up of fewer characters than the longer sentence which is from a much more advanced book in Oxford's Bookworms series (Degnan-Veness, 2008). The simplification of texts such as graded readers has generally been shown to promote reading comprehension and reading speed for L2 learners (Crossley, Yang, & McNamara, 2014). Even when simplified based primarily on intuition, difficulty in graded readers is based most directly on the number and frequency of vocabulary used within the texts, in addition to the degree of syntactic complexity and text cohesion (Claridge, 2012; Crossley, Allen, & McNamara, 2011). Furthermore, there is a well-known weak correlation between word length and word frequency within larger general corpora[1] ($r = .121$; see Piantadosi, Tily, & Gibson, 2011; Zipf, 1935). Kramer and McLean (2013) found that the difference in average word length between books read by the same student in one semester of extensive reading treatment was as large as 9.7% (see Table 1). Similarly, a difference of 12.8% was found in the average word length between two books read by students in the same class, and 18.4% was found between two books in the same extensive reading library.

While these examples were selected to illustrate the disparity, it is nevertheless apparent that the difference in the amount of text within different books can be quite large due to differences in average word length, even between books with similar running word counts.

Table 1. *Selected differences in average characters per word (cpw)*

	Title	Level		cpw	Difference in average word length (cpw)
		ERF	YL		
Individual	*Blood Diamonds*	5	1.5	5.41	9.7%
	Goodbye, Hello	2	.6	4.91	
Class	*Gulliver's Travels in Lilliput*	4	.8	5.50	12.8%
	Rain, Rain, Rain	2	.6	4.84	
Library	*Martin Luther King*	9	3.3	5.77	18.4%
	Lost at Sea	3	.9	4.80	

Note: ERF = Extensive Reading Foundation levels (found at www.erf.org); YL = *Yomiyasusa level* (a reading ease scale for graded readers); cpw = characters per word. Adapted from Kramer & McLean (2013).

Gaps in the Literature

If differences in the amount of text due only to variation in average word length were found to affect the time required for L2 readers to finish a text, it would have implications for any study which relies on the unit of whole words to measure L2 reading fluency. Such differences in text length when measured in characters, if unaccounted for, could also be a confounding variable in studies which utilize reading amount as the independent variable (i.e., time-on-task), such as in many extensive reading studies.

Some researchers (e.g., Beglar & Hunt, 2014; Beglar et al., 2012; Huffman, 2014; McLean & Rouault, 2017) have argued for the use of the standard word unit, defined as six characters including spaces and punctuation (Carver, 1976), for more precise comparisons of reading amount across studies which utilized texts of different difficulties. Such arguments have thus far been based on first language research, however, and the need for such precision has not yet been explored for those researching EFL contexts. In particular, the need for such precision in classroom-based research and pedagogical practices such as timed reading and extensive reading has not been explored.

Research Hypotheses

Two studies were conducted to investigate the effect of text length, measured in characters rather than words, on reading time measurements for Japanese learners of English. The a priori hypotheses being investigated are as follows:

1. When reading parallel passages controlled for content and word counts but differing in the number of characters they contain, the reading times of Japanese university EFL

students will be significantly longer for the passages with greater numbers of characters.

2. Throughout an intact semester-long timed reading course which has been controlled for the number of words per passage, the number of characters within each passage will significantly correlate with reading times for Japanese university EFL students.

These hypotheses, if confirmed, would support the use of character-based units when measuring reading fluency for L2 English learners.

Study 1: Average word length and L2 reading time during isolated readings

The first study looks at the effect of text length, measured in characters rather than words, on reading fluency measurements for Japanese university EFL students. To examine this effect, four reading passages were created by manipulating the average word length within each and an experimental semi-random counterbalanced Latin square design was utilized.

Participants

Japanese undergraduate students enrolled in compulsory EFL classes taught by the authors ($N =$ 212) were present for all four treatments and agreed to participate in the study. These students attended one of four private universities in western Japan and were all humanities majors from a range of English proficiency as determined by their department *hensachi*, a t-score based on the results of a standardized test taken by students in their last year of high school before entering their respective university departments. Hensachi are calculated by large university-preparation schools (i.e., cram schools) as a ranking system for Japanese universities and the departments within them. Although the scores differ slightly depending on the year they were calculated and the organization which collected the data (in this case Benesse Corporation, 2016), hensachi rankings have been shown to be a reasonable predictor of other proficiency measures including scores on the Test of English for International Communication (TOEIC) and vocabulary size estimates when direct measurements are not available (McLean, Hogg, & Kramer, 2014; Newfields, 2006). The mean hensachi score is set to 50, with 10 points above or below representing one standard deviation above or below this mean, respectively. The range of their hensachi scores was 45 to 70, indicating a large range of ability (2.5 standard deviations) within this sample relative to Japanese tertiary education as a whole. Intact classes were used, with the respective treatment groups randomly assigned within each class. After completing the readings, 41 students were removed for failing to demonstrate sufficient comprehension of the texts, leaving 171 students in the sample. The descriptive statistics are shown later in Table 4.

Materials

Adapted reading passages. Four passages were used for the primary measurements of the dependent variable, reading time. These passages consisted of two sets of two parallel adaptations: a graded reader story intended for extensive reading and a timed reading passage, two types of texts that students regularly encounter and are often used in reading research (see Appendix A for samples of text). The first set of reading passages was created from Foundation

Reading Library's *Sarah's Surprise* (Waring & Jamall, 2006). This story was chosen as it represents the narrative genre often found in books within beginning extensive reading programs and was created using highly controlled language for beginning L2 readers. In addition, ten true or false comprehension questions were created for each passage using the details and vocabulary specific to each (see Appendix B for the list of comprehension questions). True or false questions were used, rather than the standard multiple-choice questions, because of the difficulty in making suitable distractors using a story of this length (557 words for both versions, see Table 2 for more detail). Piloting confirmed that students were not able to correctly answer 70% of the questions without reading and comprehending the text.

Table 2. *Descriptive information for the texts used*

Original Passage	Version	Words	Characters	cpw	BNC/COCA 2,000 Word Family Coverage
Sarah's Surprise	Original	553	2,864	5.18	97.67 %
(Extensive Reading)	Short	557	2,723	4.89	96.89 %
	Long	557	3,112	5.59	99.27 %
"The History of the Telephone"	Original	303	1,696	5.60	99.01 %
(Timed Reading)	Short	303	1,575	5.20	99.01 %
	Long	303	1,721	5.68	99.01 %

Note. cpw = characters per word, a measure of average word length.

In order to isolate the effect of text length (measured in characters) on reading time and avoid the inclusion of possible confounds such as word frequency or grammatical difficulty, it was necessary to create sets of passages which differed only in the average length of words (measured in characters per word [cpw]), to the greatest degree possible. To accomplish this, the passages were modified by changing the proper nouns and other available words within the text to either longer or shorter forms (see Appendix C for the full list of substitutions). For the story *Sarah's Surprise*, for example, the main character Sarah became either Jennifer or Ann within the respective long and short passages. Similarly, the café name, The Lagoon, became The Sea Café or The Beach Restaurant. The detailed characteristics for each text can be seen in Table 2. The lexical coverage analysis using VocabProfile (Cobb, 2018) showed that the shorter adaptation of this text had slightly less coverage by the first 2,000 words in Nation's (2012) British National Corpus/Corpus of Contemporary American English (BNC/COCA) word family list. This difference was not considered to be problematic, however, as it was caused by words such as café and cookies. Although these words fell outside of the first 2,000 words, post-hoc interviews determined that students could correctly identify their meanings due to their use in Japanese as commonly known loanwords.

A similar procedure was used to create two adaptations for the passage titled, "The History of the Telephone" from Nation and Malarcher's *Reading for Speed and Fluency 2* (2007). This passage was chosen as it represented a different type of written genre than the graded reader, being descriptive, non-fiction, and intended for timed reading. While the passage did not have any readily available person names that could be easily adapted, a parallel form with less average characters per word was created by using alternative forms of known nouns such as phone rather

than telephone, USA rather than United States of America, and cell phone rather than mobile phone. It was also necessary to add a few additional phrases to both passages in order to maintain equivalent word counts (see Appendix C for the full list of substitutions). The properties of these passages can be seen in Table 2. Finally, ten true or false comprehension questions were created for each version (see Appendix B for the list of comprehension questions), utilizing their respective vocabulary. Piloting of these passages and comprehension questions ($N = 13$) confirmed that the passages seemed to be of equivalent difficulty and, although some students were able to guess the correct answers to some questions from prior knowledge due to the non-fiction nature of the text, reading and comprehending the passage was necessary to score above 70%, the criterion set for inclusion within the sample.

The New Vocabulary Levels Test. A 72-item version of the bilingual New Vocabulary Levels Test (NVLT) (McLean & Kramer, 2015; 2016) was administered prior to the study to confirm that the participants possessed a level of vocabulary knowledge necessary to comprehend the high-frequency English words making up the passages used in Study 1. The test consists of five 24-item levels intended to measure knowledge of English lexis from the first five 1000-word frequency levels from Nation's (2012) BNC/COCA lists along with the Academic Word List (Coxhead, 2000). As shown in Table 2, the most frequent 2,000 word families in Nation's (2012) BNC/COCA lists provide almost complete coverage of the passages used in this study. Therefore, as recommended by Nation (2012), the first three levels of the NVLT were administered (Hu & Nation, 2000; Laufer, 1989). The internal reliability of the NVLT data collected was calculated using WINSTEPS 3.73.0 (Linacre, 2013) and found to have a Rasch item reliability estimate of .97[2].

Procedure

Early piloting showed that reading two similar passages produced a sizable practice effect where students read the second passage considerably faster simply due to content familiarity. To counteract this practice effect, the students in each class were assigned to one of four groups in a semi-random counterbalanced Latin square design. It is semi-random in that although the initial treatment for each group was randomly assigned, the following treatments were then subsequently determined by this initial group placement. The counterbalanced nature of the design controls for the repetition within the repeated measures design by placing each passage in every possible position within the administration order and ensuring that each modified passage occurs before or after its associated pair an equal number of times. The Latin square design made it possible to strike a compromise between validity and practicality by testing a variety of conditions without the necessity for exhausting every option. Table 3 shows the reading order for each assigned group.

Table 3. *The semi-random counterbalanced Latin square research design*

Group	N	Week 1		Week 2	
		Reading 1	Reading 2	Reading 1	Reading 2
1	40	ER-short	TR-short	ER-long	TR-long
2	40	TR-short	ER-long	TR-long	ER-short
3	40	ER-long	TR-long	ER-short	TR-short
4	40	TR-long	ER-short	TR-short	ER-long

Note. ER = extensive reading; TR = timed reading; short = texts modified to have a shorter average word length; long = texts modified to have a longer average word length.

At the beginning of each class a display was posted for all to see with explicit instructions for the upcoming activities. First, the students completed a timed reading passage unrelated to the study, recording how long it took them to read the passage after looking at a large digital stopwatch displayed in the front of the class. The students then attempted to answer the accompanying comprehension questions without referring to the timed reading passage. Following this task, the students continued to warm up by working in pairs and reading the previous timed reading passage out loud for one minute each, with their partner monitoring their progress. After one round they repeated the reading for another minute each, trying to read further than their previous attempt. After completing these warm-up tasks, the students in each class were randomly assigned to one of the four counterbalanced groups in which they read one of the four modified passages, timing their reading using the class stopwatch displayed with the projector. After writing the time it took them to complete the reading they turned the paper over and answered the 10 comprehension questions without referring back to the original passage. The second reading passages were then distributed according to the order designated by the treatment group placement as seen in Table 3. The entire procedure, including warm-up tasks using a different timed reading passage, was repeated for Week 2 using the remaining modified passages. The teachers (the authors of this study) monitored the students closely to ensure that all procedures were followed as instructed and noticed no anomalies. Finally, in accordance with the universities' policies, the students were informed of how their data would be used and given the option to remove themselves from the study with no consequence to their grades.

Analysis

In defining the construct of reading fluency, it is necessary to include text comprehension as well as speed. Due to the importance of comprehension in defining the process of reading (Grabe, 2009; McLean, 2014), a cut-off criterion of 70% average score on the comprehension questions on all passages was set. This criterion is in accordance with the stated recommendations for timed reading programs (e.g., Quinn et al., 2007) and is a common standard for demonstrating sufficient reading comprehension (e.g., Anderson, 2008; Beglar et al., 2012; Huffman, 2014; Nuttall, 2006; McLean & Rouault, 2017). Using this criterion, 41 students were not included in the study due to insufficient evidence of text comprehension, producing a reduced sample size of 171 students.

While the Latin square design controls for a practice effect when comparing within-group measures, it is not robust against changes in the number of students within each group because its primary purpose is to counterbalance the groups. While removing participants is never ideal,

it was necessary in order to maintain equal numbers of participants between treatment groups. Thus, students whose reading times supported our hypotheses in either of the two readings (i.e., they read the shorter passage in less time) were randomly sampled and removed from the study. Although increasing the risk of Type II error, this conservative approach was preferred as it would provide stronger evidence of the hypothesized effect. In this way 11 students (see Table 5 for their descriptive statistics) were removed from the analysis, resulting in the 160 participants included in the final analysis described in Table 4.

Table 4. *Descriptive statistics for the participants who showed sufficient text comprehension*

School	*Hensachi*	N	Removed to Balance Groups	Final N	Average NVLT Scores (SD)		
					1st 1000	2nd 1000	3rd 1000
1	45-48	111	10	101	22.0 (1.8)	18.8 (3.0)	13.9 (3.4)
2	70	12	0	11	23.8 (0.5)	22.7 (0.7)	20.7 (2.0)
3	61	11	1	11	23.2 (1.3)	21.1 (1.3)	17.4 (2.5)
4	68	37	0	37	23.0 (1.0)	19.0 (3.1)	15.2 (2.4)
Total		171	11	160			

Note. Hensachi refers to a standardized *t*-score based on the results of a standardized test taken by students in their last year of high school; Average NVLT scores are for the final sample ($N = 160$) and measured by 1000-word band of Nation's (2012) BNC/COCA list – Maximum score of 24 per level.

Table 5. *Descriptive statistics for the 11 students removed to balance the Latin square subgroups*

ID	Gender	School	Treatment Group	ER (in seconds)		TR (in seconds)		Avg Score
				Short	Long	Short	Long	
27	F	1	3	120	203	155	177	9.5
54	M	1	3	208	221	123	151	8.5
55	M	1	3	88	101	84	51	7.8
110	F	1	3	158	170	82	84	9.5
117	F	1	4	96	93	48	49	9.3
134	M	1	4	293	218	219	238	8.0
175	M	1	3	184	167	58	122	8.8
178	M	1	4	302	360	240	190	8.5
187	M	1	3	230	301	221	185	9.3
190	M	1	4	232	320	212	237	9.0
243	F	2	3	92	107	51	62	9.5

Note. ER = extensive reading; TR = timed reading; Avg Score = The average of the four comprehension checks after each reading (maximum score of 10 points for each).

The independent variable was comprised of the categories short and long, representing the passages with shorter or longer text lengths when measured in characters, produced by manipulating the average word lengths. Reading time, measured in seconds, was used as the dependent variable (see Table 6 for descriptive statistics). For each set of reading passages, extensive reading and timed reading, a paired *t* test was conducted to examine whether the mean difference between the two reading times was significantly different from zero. The Bonferroni approach was taken to minimize the chances of incorrectly rejecting the null hypotheses (Type I errors) across the multiple tests. This adjustment required dividing the alpha (.05) by the number

of computed correlations, in this case two. For the paired *t* test, the *p* value must thus be less than .05 divided by 2, or .025, for the results to be considered significant. The distributions of differences in reading times were checked for normality by inspecting the descriptive statistics (see Table 6) and histograms; all distributions were judged to be sufficiently normal.

Table 6. *Descriptive statistics of the average readings time for the four reading passages (in seconds)*

Text		Min	Max	M	SEM	SD	Skewness	SES	Kurtosis	SEK
ER	Short	68	370	190.54	4.68	59.18	0.55	0.19	0.33	0.38
	Long	76	350	202.25	4.33	54.74	0.37	0.19	-0.06	0.38
	Difference	-121	165	11.71	3.81	48.22	-0.04	0.19	0.18	0.38
TR	Short	42	300	129.29	3.76	47.62	0.88	0.19	1.02	0.38
	Long	46	312	134.49	3.86	48.81	0.75	0.19	0.74	0.38
	Difference	-80	67	5.20	1.99	25.12	-0.23	0.19	0.51	0.38

Note. *N* = 160; ER = extensive reading passages; TR = timed reading passages; *SEM* = Standard Error of the Mean; *SES* = Standard Error of Skewness; *SEK* = Standard Error of Kurtosis.

Results and Discussion

Evidence of student comprehension can be seen in Table 7, which shows the descriptive statistics for the comprehension questions for each respective passage.

Table 7. *Descriptive statistics for the reading comprehension scores (K = 10)*

Text		Min	Max	M	SEM	SD	Skewness	SES	Kurtosis	SEK
ER	Short	7	10	9.48	0.06	0.70	-1.30	0.19	1.56	0.38
	Long	7	10	9.38	0.06	0.77	-1.27	0.19	1.48	0.38
TR	Short	7	10	8.76	0.08	0.96	-0.31	0.19	-0.86	0.38
	Long	7	10	8.66	0.07	0.92	-0.25	0.19	-0.72	0.38

Note. *N* = 160; *k* = 10 for each passage; ER = extensive reading text; TR = timed reading text; *SEM* = Standard Error of the Mean; *SES* = Standard Error of Skewness; *SEK* = Standard Error of Kurtosis.

Two paired-sample *t*-tests were conducted to evaluate whether the students' reading times for each set of two passages was significantly different. The results indicated that the time required to read the longer extensive reading passage (*M* = 202.25, *SD* = 54.74) was significantly longer than the time required to read the shorter version (*M* = 190.54, *SD* = 59.18), *t*(159) = 3.07, *p* = .003, *d* = .24. The 95% confidence interval for the mean difference between the two measurements was 4.18 to 19.23 seconds. Similarly, the longer version of the timed reading passage (*M* = 134.49, *SD* = 48.81) took significantly more time to read than the shorter version (*M* = 129.29, *SD* = 47.62), *t*(159) = 2.62, *p* = .010, *d* = .21. The 95% confidence interval for the mean difference between the two measurements was 1.28 to 9.12 seconds. These data indicate significant differences in the reading time between the two sets of passages. The effect sizes could be considered very small by the standards outlined by Plonsky & Oswald (2014), but this is expected for four reasons: (a) L2 reading is a multifaceted and complex process, with a wide variance in reading rates due to a number of different factors irrelevant of average word length (Jeon & Yamashita, 2014); (b) the study utilized a classroom research approach in which students recorded their own reading times, introducing more variance from measurement error above and beyond average word length; (c) the degree that we were able to edit the average length of words within the passages without changing their readability was very limited; (d) in order to maintain

equal numbers of participants in each treatment group, 11 students were removed, all with reading times supporting the hypothesis. Nevertheless, this study demonstrates that despite texts consisting of the same number of words, the average length of words in these texts was influential in determining the L2 reading times for these Japanese university EFL students.

Study 2: Text length and L2 reading time within a timed reading course

Testing the second research hypothesis, Study 2 explores to what degree the reading times of L2 learners can be predicted by the amount of text, measured in characters rather than words, in a series of intact timed reading passages. In Study 1 the average word lengths of passages were manipulated in order to produce sets of texts with shorter or longer text length as measured in characters. In this study, no manipulation was necessary so the total character counts for each intact passage were used (see Table 8 for more details).

Participants

The participants were 35 students in an intact EFL class focused on reading and listening in a university in western Japan. Only students who participated in all sessions and received sufficient scores on the comprehension questions were included in the final data set, resulting in a final sample of 27 students.

Materials

Throughout a 15-week semester the students used the first 13 passages from *Reading for Speed and Fluency*, Book 1 (Nation & Malarcher, 2007). A timer with a large, easy-to-read display was shown on the wide-screen television in the classroom. In order to determine the text characteristics for analysis, the books were scanned and the text was extracted using the optical character recognition software ABBYY (v. 12)[3]. The texts were then checked and the words and characters were counted using Microsoft Word (Windows Version, 2013). Although the texts were created to be approximately equal in length of around 300 words, in fact, they ranged from 295 to 305 words, an issue raised by Tabata-Sandom (2017) as a possible explanation for differences in reading times.

Procedure

The timed reading passages were used as in-class timed reading practice, administered from the second week of class until week 14. Each week the students would first complete a weekly vocabulary quiz and 30 minutes of sustained silent reading practice using in-class graded readers before the timed reading activity. In accordance with the university's policies, the students were informed of how their data would be used and given the option to remove themselves from the study with no consequence to their grades. The students were told the purposes of timed reading practice and tried to read faster each week. They were instructed to read each passage as quickly as possible, writing their completion time in their books upon finishing. After reading, the students answered five multiple-choice comprehension questions. In contrast with Study 1, in which ten original comprehension questions were created for each passage, this study simply

used the five questions included with each passage, in a manner consistent with the textbook's intended use.

Analysis

In order to conduct this analysis, it was necessary to ensure that the students read with adequate comprehension (see McLean, 2014). For this reason, all students with average reading comprehension scores below 70% were removed from the study ($n = 3$). An additional 5 students were removed from the analysis due to absence from class, reducing the sample size from 35 to 27. Before conducting the analysis, the descriptive statistics and histograms of reading times for each passage were visually inspected and judged to be sufficiently normal. A two-tailed Pearson's partial correlation was calculated between the number of characters in each passage and the average time required to read the same passage, controlling for the number of words. This partial correlation is preferable over a standard correlation because the variance in numbers of words in each passage could reasonably be presumed to be the common cause of any correlation found between character counts and reading times, as suggested by Tabata-Sandom (2017). In such a scenario, we would expect the correlation between these two variables partialling out the effects of word counts to be approximately equal to zero.

Results and Discussion

Table 8 shows the average reading time for each timed reading passage, as well as the word and character counts used for the analysis.

Table 8. *Average time for each timed reading passage*

Passage	Words	Characters	Min	Max	*M*	*SEM*	*SD*	Skewness	*SES*	Kurtosis	*SEK*
1	305	1636	60	210	141.07	7.03	36.54	0.13	0.45	0.11	0.87
2	299	1610	54	192	128.22	5.49	28.53	-0.39	0.45	1.13	0.87
3	301	1595	73	250	161.81	7.30	37.93	0.08	0.45	0.18	0.87
4	300	1709	77	250	149.37	6.85	35.60	0.46	0.45	1.55	0.87
5	296	1474	44	210	138.67	6.37	33.12	-0.61	0.45	1.58	0.87
6	301	1522	47	205	144.15	6.76	35.11	-0.64	0.45	0.84	0.87
7	303	1465	51	195	136.63	5.98	31.10	-0.42	0.45	1.18	0.87
8	296	1566	39	232	137.74	7.05	36.63	0.00	0.45	1.99	0.87
9	295	1391	38	179	113.07	5.92	30.78	0.13	0.45	0.67	0.87
10	302	1699	35	222	153.70	7.20	37.39	-0.93	0.45	2.97	0.87
11	304	1492	43	230	123.19	6.07	31.55	1.02	0.45	5.46	0.87
12	300	1607	43	196	137.96	5.78	30.03	-1.10	0.45	3.14	0.87
13	299	1546	52	162	112.22	4.80	24.93	-0.34	0.45	0.30	0.87

Note: N = 27; *SD* = Standard Deviation; *SEM* = Standard Error of the Mean; *SES* = Standard Error of Skewness; *SEK* = Standard Error of Kurtosis.

A partial correlation coefficient was computed between the average reading time and number of characters within each passage, holding constant the number of words. The correlation, $r = .56$, $p = .061$, was not significant at the .05 alpha level, but could be considered to have a large effect

size according to Plonsky & Oswald's criteria for effect size interpretation (2014). What this indicates is that the number of characters within each passage showed a strong relationship with the average time required to read that passage, controlling for the small differences in word counts to test the claim by Tabata-Sandom (2017).

For a visual representation of this relationship see Figure 2, which shows the predicted reading times of a theoretical student who read the first 13 of the timed reading passages in Nation and Malarcher (2007) at a constant reading rate *measured in characters per minute* (shown with the dashed line) compared with the actual average student reading time from this study ($N = 27$) for the same passages (shown with the solid line). Although the actual student data differ from the predicted reading times for the first few treatments, after the students acclimated to the task the average reading times demonstrated the same patterns as those predicted by the amount of characters within each passage. These findings suggest that for classroom reading fluency treatments, teachers, administrators, and researchers should consider the use of character-based measurement to more accurately interpret their data.
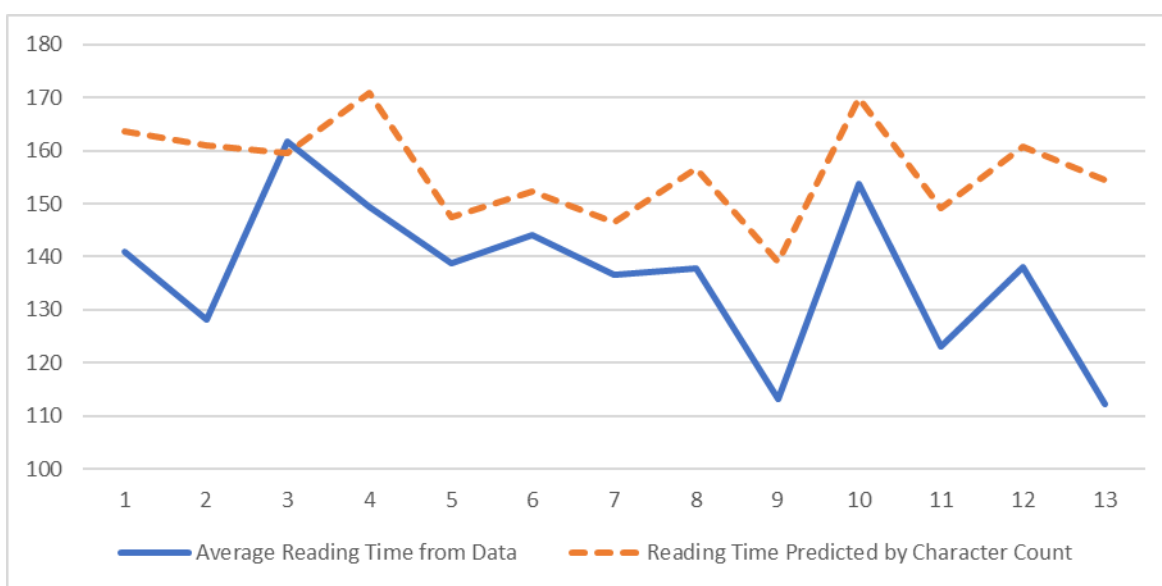

Figure 2. Average student reading time ($N = 27$) in seconds (solid line) and the predicted reading times (dashed line) of a theoretical student who reads every passage at a constant rate of 600 characters per minute.

Overall Discussion and Implications

The results of the first and second studies together demonstrate that the number of characters within a reading passage has a measurable effect on the time required for Japanese learners of English to read that passage which is not explained by simply counting the number of words. This finding is important because the reading times and reading rates reported in published research are influenced by not only the number of words within a text, but also the average length of those words. While the results for the first study showed that this difference in reading time can be induced by artificially manipulating the average word length for isolated passages, the results for the second study suggest that this effect manifests even among completely

different passages within a timed reading course. If care is not taken to account for text length as measured in characters, researchers and teachers alike might interpret reading time variation in ways which are not appropriate for the target construct being measured, namely reading fluency.

Given the results of these two studies, there are two implications that should be discussed. First, we know that the amount of time it takes an L2 learner to read particular texts can differ significantly depending on a variety of factors, including the average word length or total number of characters. Measured reading times or reading rates should not be considered a fixed reader trait, but rather the result of the interaction between the reader, the context, and the specific text being read. Furthermore, any generalizations made from these measurements should be done with care. This has direct pedagogical implications, especially for teachers utilizing timed reading treatments in their classrooms. Seemingly sudden increases in the time required to read certain passages may simply be a result of longer words within that passage and may not be a cause for concern. Letting the students know about such patterns might also help them avoid negative feelings if they fail to maintain or improve their reading times each week.

A second implication of these results relates to the use of inferential statistical techniques such as the ubiquitous Student's *t* test or Analysis of Variance (ANOVA) within reading research. These statistical tests require an often-ignored assumption of at least interval scale measurement. The defining property of interval scale measurement is that one unit is equivalent to all others; in other words, the difference between 10 and 20 units is equivalent to the difference between 50 and 60 units (Field, 2009, p. 359). As the studies presented in this paper have demonstrated, one word unit is often not equivalent to others, but can instead be thought to represent words which contain both many and few characters. As shown in this study, the number of characters within a word has a measurable effect on reading times for Japanese learners of English, violating the assumption of interval scale measurement for the word unit. Measuring text amounts or reading rates using character-based units would address this limitation.

One method of addressing the described variation in word length throughout various texts is to use Carver's (1976) recommended unit of text amount, the *standard word*. As Carver stated referring to L1 reading measurement, "any investigation of reading rate which purports to be precise should control for word length, especially when the material is at different grade difficulties" (p. 197). We can see this unit of measurement in use in limited L2 studies (e.g., Beglar & Hunt, 2014; Beglar et al., 2012; Huffman, 2014; McLean & Rouault, 2017; Shimono, 2018), although the need for such a measurement unit in L2 reading research has not been clearly demonstrated until now. A weakness to exclusively using the standard word unit, however, is that information regarding the number of running words within a text is lost. An alternative approach to the standard word unit would be to report not only the word count for each reading passage, but also the average word length in characters per word (cpw) or the total number of characters along with other measurements such as vocabulary frequency analyses that might affect a student's reading rate. This would allow researchers to calculate the number of standard words if necessary and gives the maximum flexibility to readers for interpreting study results. If necessary, an updated list of word and character counts for all known timed reading passages can be found at the first author's personal site (https://brandonkramer.net/resources).

In summary, researchers should utilize measurement units that maximize accuracy and precision in a way that is construct relevant. While simple word counts can be considered appropriate when investigating word recognition and activation at the whole-word level, when considering the automatization of sub-lexical processing the raw amount of text in characters is also an important consideration and worth measuring accurately and precisely, especially given its effect on the L2 reading time of Japanese university EFL students as demonstrated in these two studies.

Limitations and Conclusion

Firstly, the research design was limited in several ways by its classroom nature, especially when compared with more controlled laboratory experiments. Despite lacking a degree of methodological purity, however, this study shows greater ecological validity by showing that word length has a measurable effect on reading time in nonlaboratory classroom contexts (see Hudson & Llosa, 2015). Furthermore, the lengths of the passages in Study 1 were manipulated so that the syntax and lexical coverage were nearly identical, but it is conceivable that this introduced other confounds such as differences in name familiarity. Further research could expand upon this by controlling for such variables within the passages and providing more evidence that the passages were truly equivalent except for the number of characters. Finally, although the correlation in Study 2 was close to large by Plonsky & Oswald's (2014) criterion, the results lacked statistical significance at the .05 level. While it is possible that including more passages in the data would resolve this issue, further research should be carried out to bolster the generalizability of these claims.

Despite these limitations, the theory-driven hypotheses were largely supported and the effects of word length on reading times in these studies were clear. Overall, the results of these two studies highlight the need for using character-based units when measuring reading fluency and including descriptive qualities of texts such as average word length or total character counts within future L2 reading research. If researchers work to increase the accuracy and precision of their measurements, the strength of their inferences and claims will also grow, helping the field to move forward.

Notes

1. Calculations of correlations between word frequency and word length actually use the logarithmic transformation of frequency due to the severe skewness of the data.

2. Rasch reliability estimates can be considered equivalent in interpretation to Cronbach alpha statistics, although in reality they are slightly more conservative and preferable because they are based on data which stochastically conform to the Rasch model.

3. To obtain the word and character counts used for the analyses, a digital version of the text was necessary. Scanning pages typically only produces an image file, which is simply a picture of the words on the page. Using a program such as ABBYY (v. 12), the computer can automatically recognize the contents of the image as text on a page and store it digitally in text format.

Acknowledgements

References

Anderson, N. J. (2008). *Practical English language teaching: Reading*. New York, NY: McGraw-Hill.

Axten, S. (2008). *Blue fins*. London, UK: Macmillan Education.

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

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

Benesse Corporation. (2016). Kiboudai kitto mitsukaru kensaku [Search to find the university you hope for]. Retrieved July 4, 2016, from https://kou.benesse.co.jp/hensachi/index.php

Carver, R. P. (1976). Word length, prose difficulty, and reading rate. *Journal of Literacy Research*, *8*, 193–203. doi:10.1080/10862967609547176

Chang, A., & Millett, S. (2013). Improving reading rates and comprehension through timed repeated reading. *Reading in a Foreign Language*, *25*, 126–148.

Claridge, G. (2012). Graded readers: How the publishers make the grade. *Reading in a Foreign Langauge*, *24*, 106–119.

Cobb, T. (2018). *Compleat web VP* [web application]. Accessed 2 May 2018 at http://www.lextutor.ca/vp/comp/

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge: Cambridge University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238. doi:10.2307/3587951

Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language, 23*(1), 84–101. Retrieved from http://nflrc.hawaii.edu/rfl/April2011/articles/crossley.pdf

Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language, 26*, 92–113. Retrieved from http://nflrc.hawaii.edu/rfl/April2014/articles/crossley.pdf

Degnan-Veness, C. (2008). *Martin Luther King*. Harlow, England: Pearson Education.

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

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge: Cambridge University Press.

Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd ed.). New York, NY: Routledge.

Hitosugi, C. I., & Day, R. R. (2004). Extensive reading in Japanese. *Reading in a Foreign Language*, *16*, 20–39.

Horiba, Y., & Fukaya, K. (2015). Reading and learning from L2 text: Effects of reading goal, topic familiarity, and language proficiency. *Reading in a Foreign Language*, *27*, 22–46.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430.

Hudson, T., & Llosa, L. (2015). Design issues and inference in experimental L2 research: Design issues and inference. *Language Learning, 65*(S1), 76–96. doi:10.1111/lang.12113

Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language*, *26*, 17–33.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*, 160–212. doi:10.1111/lang.12034

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York, NY: Routledge.

Kramer, B., & McLean, S. (2013). Improved reading measurement utilizing the standard word unit. In S. Miles & M. Brierley (Eds.), *Second World Congress in Extensive Reading Proceedings* (Vol. 2, pp. 117–125). Seoul, Korea: Extensive Reading Foundation. Retrieved from http://erfoundation.org/ERWC2-Proceedings.pdf

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. doi:10.1016/0010-0285(74)90015-2

Lake, J. W. (2014). Curious readers and interesting reads: Developing a positive L2 reading self and motivation through extensive reading. *Journal of Extensive Reading*, *2*, 13–27.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Clevedon, England: Multilingual Matters.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*, 15–30.

Linacre, J. M. (2013). *A user's guide to WINSTEPS*. Chicago, IL: MESA. Retrieved from winsteps.com

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading and Writing Quarterly*, *13*, 123–146. doi:10.1080/1057356970130203

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*, 861–904. doi:10.1017/S0142716418000036

McLean, S. (2014). Addressing the importance of comprehension to reading: Learning lessons from Chang (2012). *Reading in a Foreign Language*, *26*, 186–191.

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. doi:10.7820/vli.v03.2.mclean.et.al

McLean, S., & Kramer, B. (2015). The creation of a new vocabulary levels test. *Shiken*, *19*(2), 1–11.

McLean, S., & Kramer, B. (2016). The development of a Japanese bilingual version of the New Vocabulary Levels Test. *Vocabulary Education and Research Bulletin*, *5*(1), 2–5.

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

Millet, S. (2005). *New Zealand speed readings for ESL learners: Book one*. Wellington: English Language Institute Occasional Publication.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Rowley, MA: Newbury House.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Nation, P. (2012). *The BNC/COCA word family lists*. Unpublished paper. Retrieved from victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf

Nation, P., & Malarcher, C. (2007). *Reading for speed and fluency*. Tokyo: Compass.

Newfields, T. (2006). Assessment literacy self-study quiz #1 [Suggested answers]. *Shiken*, *10*(2), 25–32.

Nuttall, C. (2006). *Teaching reading skills in a foreign language* (2nd ed.). Oxford: Heinemann Educational Books.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences, 108*, 3526–3529. doi:10.1073/pnas.1012551108

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*, 878–912.

Quinn, E., Nation, I. S. P., & Millet, S. (2007). *Asian and Pacific speed readings for ESL learners.* Wellington: English Language Institute Occasional Publication.

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

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*, 55–88. doi:10.1177/026553220101800103

Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 382–408). Malden, MA: Blackwell.

Shimono, T. R. (2018). L2 reading fluency progression using timed reading and repeated oral reading. *Reading in a Foreign Language, 30*, 152–179. Retrieved from http://nflrc.hawaii.edu/rfl/April2018/articles/shimono.pdf

Tabata-Sandom, M. (2017). L2 Japanese learners' responses to translation, speed reading, and "pleasure reading" as a form of extensive reading. *Reading in a Foreign Language*, *29*, 113–132.

Tran, T. N. Y. (2012). The effects of a speed reading course and speed transfer to other types of texts. *RELC Journal*, *43*, 23–37. doi:10.1177/0033688212439996

Tran, T. N. Y., & Nation, P. (2014). Reading speed improvement in a speed reading course and its effect on language memory span. *Electronic Journal of Foreign Language Teaching*, *11*, 5–20.

Waring, R., & Jamall, M. (2006). *Sarah's surprise*. Independence, KY: Heinle Cengage.

Winke, P., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue: Eye-movement recordings in second language research. *Studies in Second Language Acquisition*, *35*, 205–212. doi:10.1017/S027226311200085X

Yamashita, J. (2004). Reading attitudes in L1 and L2, and their influence on L2 extensive reading. *Reading in a Foreign Language*, *16*, 1–19.

Zipf, G. K. (1935). *The psycho-biology of language* (Vol. ix). Oxford, England: Houghton, Mifflin.

Appendix A

Sample text from the adapted reading passages used in Study 1

Extensive Reading Passages

Short	Long
"Hi, Ann, I'm Mrs. Bond," says a lady.	"Hello, Jennifer, I'm Mrs. Harrison," says a woman.
"How do you do, Mrs. Bond?" says Ann.	"How do you do, Mrs. Harrison?" says Jennifer.
Mrs. Bond has a cafe. The cafe's name is The Sea Cafe.	Mrs. Harrison has a restaurant. The restaurant's name is The Beach Restaurant.

Both versions are adapted from the original text found in *Sarah's Surprise*, from the Foundation Reading Library series:
Waring, R., & Jamall, M. (2006). *Sarah's surprise*. Independence, KY: Heinle Cengage.

Timed Reading Passages

Short	Long
About 130 years ago, people began to use phones in their homes. The first phones had one part that people held to their ears and one part they talked into. There were no numbers on these phones. People just talked to anyone who had a phone. The big problem was that it cost a lot to have and use a phone so not many people had them.	About 130 years ago, people started to use telephones in their houses. The first telephones had one part that people held to their ears and another part they talked into. There were no numbers on these telephones. People just talked to anyone who had a telephone. The large problem was that it was expensive to have and use a telephone, so not many people owned them.

Both versions are adapted from the original text found in *"The History of the Telephone"* found in:
Nation, P., & Malarcher, C. (2007). *Reading for speed and fluency*. Tokyo: Compass.

Appendix B

Comprehension questions for the extensive reading passages used in Study 1

Short
Questions: True or False.

1. The name of the restaurant is Jolly pasta.	True / False
2. Ann has worked at the shop for one month.	True / False
3. Mrs. Bond goes away from the cafe to go shopping.	True / False
4. The man takes a cake from the cafe.	True / False
5. The man gives Ann money.	True / False
6. Ann calls the police after the man leaves.	True / False
7. The man is wearing black jeans and a black shirt.	True / False
8. Mrs. Bond is surprised when Ann tells her what she saw.	True / False
9. The man often gets sandwiches from the cafe.	True / False
10. The man is married to Mrs. Bond.	True / False

Long
Questions: True or False.

1. The name of the restaurant is Jolly pasta.	True / False
2. Jennifer has worked at the shop for one month.	True / False
3. Mrs. Harrison goes away from the restaurant to go shopping.	True / False
4. The man takes a cake from the restaurant.	True / False
5. The man gives Jennifer money.	True / False
6. Jennifer calls the police after the man leaves.	True / False
7. The man is wearing black jeans and a black shirt.	True / False
8. Mrs. Harrison is surprised when Jennifer tells her what she saw.	True / False
9. The man often gets sandwiches from the restaurant.	True / False
10. The man is married to Mrs. Harrison.	True / False

Comprehension questions for the timed reading passages used in Study 1

Short
Questions: True or False.

1. About 130 years ago all houses had phones.	True / False
2. People had to tell the operator the address of the person they wanted to call.	True / False
3. With operators, calls were made very quickly.	True / False
4. People began to buy phones because they became cheaper.	True / False
5. Recently people want more old style phones in their houses.	True / False
6. Today most people say they do not need a cell phone.	True / False
7. The first phones had many parts.	True / False
8. The first phones did not have numbers on them.	True / False
9. The first phones were very expensive.	True / False
10. By 1970 almost every house in the USA had a phone.	True / False

Long

Questions: True or False.

1. About 130 years ago all houses had telephones.	True / False
2. People had to tell the operator the address of the person they wanted to call.	True / False
3. With operators, calls were made very quickly.	True / False
4. People began to buy telephones because they became cheaper.	True / False
5. Recently people want more old style telephones in their houses.	True / False
6. Today most people say they do not need a mobile phone.	True / False
7. The first telephones had many parts.	True / False
8. The first telephones did not have numbers on them.	True / False
9. The first telephones were very expensive.	True / False
10. By 1970 almost every house in the USA had a telephone.	True / False

Appendix C

Words which were changed to create parallel versions of each text

Sarah's Surprise Adaptation		Telephone Adaptation	
Short	Long	Short	Long
Ann	Jennifer	Phone	Telephone
Hi	Hello	began	started
Bond	Harrison	homes	houses
lady	woman	one	another
café	restaurant	big	large
Sea Café	Beach Restaurant	cost a lot	was expensive
Rob	Alberto	had	owned
too	also	often	sometimes
blue	yellow	make	connect
		USA	United States of America
		way	system
		cheaper	less expensive
		them	telephones
		almost	nearly
		USA	United States
		cell	mobile
		any place	anywhere
		cell phones	mobile phone services

Phrases present in only one passage without substitutions

around the world	with operators
the year	

Note. Not all pairs of words in the timed reading passage are one-to-one substitutions

About the Authors

Brandon Kramer works, lives, and studies in Osaka, Japan. He holds an M.S.Ed. degree in TESOL from Temple University where he is currently working on his Ph.D. in Education (Applied Linguistics). His research interests fall in the intersection between vocabulary acquisition, language testing, and corpus linguistics. E-mail: brandon.L.kramer@gmail.com

Stuart McLean holds an M.S.Ed. (TESOL), a Ph.D. in Forensic Medicine, and a P.G.C.E. He is currently an Applied Linguistics doctoral student at Kansai University. He has published in *Reading in a Foreign Language, Vocabulary Learning and Instruction, Language Teaching Research, TESOL Quarterly, Applied Linguistics, System, and Language Assessment Quarterly*, on the topics of language assessment, research methodology, vocabulary learning, corpus linguistics, extensive reading, TESOL, and listening. E-mail: stumc93@gmail.com