Reading in a Foreign Language April 2016, Volume 28, No. 1
ISSN 1539-0578 pp. 118–142

Measuring second language vocabulary knowledge using a temporal method

Masayuki Tanabe
Aoyama Gakuin University
Japan

Abstract

The present study addressed the role of speed as a factor in tests of second language (L2) vocabulary knowledge, presupposing that speed of performance is important in actual language use. Research questions were: (a) Do learners with a larger vocabulary size answer faster on an L2 vocabulary breadth test than smaller vocabulary sized learners?; (b) Are there systematic increases in response time (RT) as word frequency decreases in an L2 vocabulary breadth test?; and (c) Do RTs of correct responses on an L2 vocabulary breadth test predict accurate and quick L2 reading? Participants were 24 Japanese university students. Results indicated that (a) vocabulary size facilitated lexical accessibility, (b) high frequency words were accessed more quickly but this was only observable after reaching a certain threshold of vocabulary size, and (c) vocabulary score (accuracy) alone was not associated with accurate and quick reading but vocabulary RT (accuracy + speed) was.

Keywords: vocabulary test, vocabulary breadth, response time, reading speed, Japanese EFL learners

Traditionally, vocabulary tests have been common tools used to assess learners' lexical knowledge in educational settings, but it is uncommon that speed of response is measured on them. However, speed is an important aspect in the sense that automatized processing distinguishes fluent from nonfluent abilities (Segalowitz & Hulstijn, 2005), and our goal in second language (L2) learning is to make L2 knowledge practically available.

Empirical studies that have looked into reaction time (RT) on an L2 vocabulary test are limited (Harrington, 2006; Laufer & Nation, 2001; Miralpeix & Meara, 2014; Pellicer-Sánchez & Schmitt, 2012). The previous studies overall suggested two major effects: (a) speed of lexical access improves as a function of increasing size of vocabulary, which is the size effect; and (b) this happens as a function of word frequency, which is the frequency effect. These studies demonstrated the relationship between breadth of vocabulary knowledge and speed of lexical access within the framework of vocabulary testing. However, what we also need to know is whether the performance measured in testing situations can be transferred to practical use of the language. In this respect, positive interrelations between vocabulary test scores and levels of reading comprehension have been widely discussed in L2 vocabulary studies (Anderson & Freebody, 1981, 1983; Beglar & Hunt, 1999; Laufer & Ravenhorst-Kalovski, 2010; Moinzadeh

& Moslehpour, 2012; Qian, 1999; Yalin & Wei, 2011; Yildirim, Yildiz, & Ates, 2011) but they have not focused on speed. Therefore, further investigation in this line of research is necessary in terms of how quickly learners respond to vocabulary questions.

Background

Measuring RTs on a Vocabulary Test

In studies of language acquisition, timed lexical recognition has been a common experimental methodology to assess automaticity of lexical access (Hulstijn, van Gelderen, & Schoonen, 2009). This lab-based task involves a forced yes-no decision of a real- or non-word presented without context (Harrington, 2006; Jiang, 2012) and thus results in faster and less variable RTs (Harrington, 2006). Therefore, it is not surprising that the majority of the previous studies dealt with the yes-no vocabulary format.

With a computerized yes-no test, Harrington (2006) compared accuracy, RTs, and degree of automaticity, operationalized as the coefficient of variation (CV) of response time (Segalowitz & Segalowitz, 1993) of intermediate and advanced learners of English and native speakers across four word-frequency bands—2,000-, 3,000-, 5,000-, and 10,000-word levels—based on the Vocabulary Levels Test (Nation, 1990). Harrington found that RT measures systematically decreased as a function of increasing proficiency. Also, response variability, as measured by CV, decreased as performance improves.

Here the CV, employed in the study above, is an index of word recognition automaticity that was proposed by Segalowitz and his colleagues (Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz, & Wood, 1998). The CV, calculated by the standard deviation of RT divided by the mean RT, reflects overall variability of the RTs in a set of trials. Responses under development in early stages of learning become faster as a function of practice. In such cases, the coefficient of variability of RTs (CV_{RT}) is not reduced whereas the mean RT becomes smaller because RTs are still variable. In contrast, CV_{RT} in later stages of learning must be reduced because RTs have become difficult to improve and thus converged on a certain level. In this way smaller mean RT together with smaller CV_{RT} provides an indication of qualitative change of the responses, or automaticity formation.

In contrast, Miralpeix and Meara (2014), who performed a yes-no vocabulary test based on the JACET List of 8000 Basic Words (JACET8000; JACET Committee of Basic Words Revision, 2003) to measure vocabulary size and a timed animate-inanimate decision task devised by Segalowitz (2003) to measure RTs showed that there were no significant systematic interrelations of vocabulary size with either RT or CV. However, because the mean RT of the group with the largest vocabulary was the fastest among the groups and the mean CVs of this group was the lowest, the authors concluded that accessibility might not play a determinant role until a bigger lexicon is acquired or that the relationship between these variables should be examined at particular stages of the learners' development.

Pellicer-Sánchez and Schmitt (2012) attempted to find whether RTs could be a major determinant of scoring for yes-no vocabulary tests instead of traditional formula-based scoring. Their results suggested that the RT approach was most appropriate for use with native speakers or advanced learners who were rarely puzzled by pseudowords and that traditional non-word formulae were still advantageous to the judgment of intermediate L2 learners. As a result, this posed a potential challenge for the yes-no test to be the best format to measure RTs of L2 learners, providing empirical support for the view that the yes-no test "does not require the learners to provide any independent evidence that they know the words they have ticked" (Read, 1988, p. 16).

An example of measuring lexical-semantic access was Laufer and Nation (2001), who employed a computerized Vocabulary Levels Test as an RT measure. They found that non-native speakers' increase in speed lags behind increase in vocabulary size. Significant fast responses to 3,000-word level words were observable in 5,000-word level learners but it was not the case for learners whose vocabulary size was less than 5,000.

Laufer and Nation's (2001) study was criticized for the use of a multiple definition-word association format, three definitions versus six words, as an RT measure because RT values and variability tend to be higher on such a format (Harrington, 2006). However, the use of the definition-word association format could be in some degree justified in the sense that semantic access of the tested words is objectively confirmable. Also large RT values are considered not to be crucial because RTs are basically relative values since any RT measures do not capture the exact elapsed time in our brains. However, as Harrington argued, their format seemed to have a variability problem in which the format assessed three target words at one question. As an RT measure, it is a serious problem that the RT for a certain word is difficult to identify. In their format, RTs for the second and the third matching of a question are not reliable because the test-taker has already scanned all option words including those target words to make the first matching. A possible solution for this is thus to set up only one target word in a question, which involves the use of a simple multiple-choice format.

Speed of Word Recognition as a Subcomponent of Reading Rate

Reading requires simultaneous mental operations and interactions between lower and higher order processing (Just & Carpenter, 1992) but the present study focused in particular on speed of lexical access. In terms of speed, whether vocabulary knowledge facilitates reading performance has been investigated in L2 reading studies. A few empirical studies dealing with this particular issue are found in relatively recent papers.

Van Gelderen et al. (2004) examined contributive subcomponents of L2 reading comprehension for Dutch students learning English in secondary education. The components included L2 word recognition and L2 sentence verification measures as speed components. Results demonstrated that the speed components lacked statistical significance in the explanation of reading comprehension. This study regarded reading comprehension as a dependent variable but the relationship between word recognition and sentence verification was not discussed. However, the correlation coefficient between those two subcomponents was moderately high (r = .61). It seems to me that this implied that speed of word recognition and reading speed were interrelated because the sentence verification task was to read short sentences and decide whether they made sense or not as fast as possible.

In a study of Dutch students aged 13 to 14, Fukkink, Hulstijn, and Simis (2005) investigated whether increasing level of automaticity of lexical recognition by training enhances reading speed and comprehension in L2. The training focused on recognition of word forms but also semantic aspects of words in which cloze exercises to make sensible sentences or translation tasks were repeatedly given. Reading speed was assessed by measuring RTs of four short paragraph readings, each followed by a simple yes-no factual question. Reading comprehension was measured by two English articles that required time-consuming thought processes. However, results showed no training effects for either reading speed or comprehension, while improvements of lexical access were clearly detected.

In contrast, Yamashita (2013) studied Japanese university students who learned English via formal education and found that speed of meaning-oriented word recognition was significantly interrelated with reading rate. The timed lexical recognition task was to make a dichotomous judgment on whether a pair of words had an antonymous relation and participants were required to complete as many judgments as possible for one minute. She argued that the semantic aspect of processing needs to be included as L2 word recognition skills for reading ability because sufficient activation of the lexical meaning in L2 is not guaranteed unlike in first language (L1) processing.

The latter two studies reviewed above emphasized semantic access in word recognition. In this respect, studies of L2 reading and those of L2 vocabulary testing share a common interest in which semantic aspects must be considered during tests of L2 word recognition. Since reading proceeds with comprehension of sentences, and a sentence consists of words, reading involves understanding the sum of the meanings of the words contained in the sentence. It would therefore be beneficial for both areas to investigate the relationship between RTs on an L2 vocabulary test involving word meanings and speed of L2 reading.

The Present Study

Research Questions

In contrast to breadth and depth of word knowledge which have been widely discussed in the field of L2 vocabulary testing including their practical influence on reading comprehension, the number of studies looking into the accessibility of lexical knowledge is still limited. Moreover, the lexical accessibility measured by timed vocabulary tests and its association with L2 reading performance have not yet been investigated. Although some L2 reading studies that investigate factors contributing to proficient L2 reading have treated lexical accessibility in terms of word recognition skill (Fukkink et al., 2005; van Gelderen et al., 2004; Yamashita, 2013), whether word recognition predicts reading performance has not been resolved (Nassaji, 2014). Yamashita (2013), who found a significant positive relationship between word recognition skills and reading performance, mentioned that “the semantic component in word recognition would deserve more attention” (p. 65) for the reason that the semantic activation in an L2 is not always sufficient for learners in contrast with automatic semantic access in their L1 because connections between L2 word forms and meanings are assumed to be weaker than the form-meaning links in L1. In this respect, learners' L2 vocabulary knowledge should be measured by a test that requires learners to demonstrate knowledge of word meanings. Therefore, the present study measured accessibility of L2 vocabulary knowledge by using a testing format that asked about the meaning of words explicitly and investigated its contribution to L2 reading performance.

The researcher addressed the three research questions shown below (RQs). RQ1 and RQ2 are partial replications of the vocabulary RT studies which we have reviewed in the earlier section. The previous studies have employed either yes-no tests or the Vocabulary Levels Test. However, both tests have shortcomings. The yes-no test does not require the test-taker to show evidence for their semantic knowledge of the words (Read, 1988). The Vocabulary Levels Test does ask for knowledge of the word meanings but its format requires the test-taker to scan the same words repeatedly, and is therefore inadequate as an RT measure. Accordingly, the present replication involves a simple multiple-choice format for matching words with their definitions. RQ3 is an extended question to examine the relationship between speed of lexical access measured on this format and reading performance.

1. Do learners with a larger vocabulary size answer faster on an L2 vocabulary breadth test than learners with smaller vocabulary sizes?

2. Are there systematic increases in RTs as word frequency decreases in an L2 vocabulary breadth test?

3: Do RTs of correct responses on an L2 vocabulary breadth test predict accurate and quick L2 reading?

Participants

The present study tested adult Japanese speakers who have learnt English primarily in the standard educational situation in Japan, assuming that such learners have the same age of onset of English learning. Consequently, participants met the following conditions: he or she (a) is a native speaker of Japanese; (b) has parents who speak Japanese and not English; (c) is not a returnee from abroad; and (d) has been educated in Japanese middle and high schools for three years each.

As a result, 24 Japanese university students—aged 20–26 (mean = 22, standard deviation (*SD*) = 1.76), five males and 19 females—participated in the study. Most of them were majoring in English literature. Others included students majoring in education and law. About a month before the current session, they had taken the TOEFL-ITP as a requirement of their university and the mean score was 497.97 (*SD* = 39.27). This score corresponded to the B1 level (Independent User – Threshold) in the Common European Framework of Reference for Languages (Educational Testing Service, 2015). They were all undergraduate students. They received 2,000 yen for their participation.

Materials

Two computer-based tests were developed for the present study. They were stand-alone software written in Microsoft Visual Basic 2010 and were executable on windows-based computers. The researcher developed them on a Windows laptop computer and used the same machine in the experiment sessions. The researcher also made a response pad for the vocabulary test by modifying a numeric keypad.

The vocabulary breadth test. The vocabulary breadth test employed a form-meaning association format following the claim by Pellicer-Sánchez and Schmitt (2012) and Yamashita (2013) that measuring the L2 lexicon in this way is important in the sense that semantic access of L2 words is not always sufficiently automatized, in contrast to access of L1 words. Each question consisted of a word definition and four option words, in which one of the option words corresponded to the definition and the other three words were distractors. The number of options was determined referring to Currie and Chiramanee (2010), who found that their participants' behavior "in the four- and five-option items was almost identical, while the only effect the three-option format had overall was to render the items slightly easier" (p. 486). The present test involved the use of an external five-button response pad. Four of those buttons were for answering and one was for skipping the question if the test-taker did not know the answer.

Each question started with presenting the option words first but the definition was not displayed. This state was kept for five seconds so that the test-taker could check the presented words during this period. Five seconds later, the definition was additionally displayed and the internal stopwatch started up to measure RT. The test-taker responded as fast as possible by pressing one of the four buttons corresponding to the word he or she thought was right, or the "pass" button when feeling the question was too difficult to answer. The presentation procedure is illustrated in Appendix A.

Theoretically, the development of learners' semantic processing of L2 words starts with associating those words with their equivalent L1 translations. The links between those L2 words and their concepts that were established via L1 translations become stronger as the learners experience similar examples, and finally the L1 translations are bypassed and direct links between the L2 words and their concepts are formed (Jiang, 2000). From this viewpoint, it is important to compose vocabulary tests in which, for example, both the definition and the option words are presented in the target language, rather than presenting either of them in their L1. This procedure was followed in the present test. In contrast, L2-L1 matching tests, as opposed to monolingual tests, are less appropriate indicators of learners' developed L2 processing because they conflate L1 and L2 processing.

The test had 63 questions in total. Thirteen of them were for the practice session and 50 of them were for the testing session. Of the 63 questions, 80% of them were English vocabulary questions and 20% were Japanese vocabulary questions. The presentation of first language questions was to validate reliability of test-takers' responses of the L2 questions. If the test-taker responded to an English question faster than his or her fastest response in Japanese questions, the answer was excluded from the analysis because such responses were not reliable. Laufer and Nation (2001) found that native speakers' responses were homogeneous across word frequencies whereas non-native speakers responded more slowly to less frequent words. In their analysis, they eliminated participants who responded very inaccurately and faster than native speakers, suspecting that those participants were random key hitters. Hulstijn, van Gelderen and Schoonen (2009, Study 1) employed the fastest RT of six advanced L2 speakers as their too-fast baseline and excluded participants' responses faster than the fastest possible reactions of the experts as outliers, reasoning that "it was considered highly unlikely that such fast reactions were based on genuine processing of the items" (p. 566). Referring to these previous examples, the present study assumed that the learner's fastest possible word access must be seen in their L1.

The English items were taken from the Vocabulary Levels Test (Schmitt, 2000; Nation, 2001). Japanese items were taken from *Japanese-Language Proficiency Test Target 2,000: N1 Vocabulary* (Obunsha, 2011). This word list contains Japanese words that may appear in the most proficient level, the N1 level, of the Japanese-Language Proficiency Test for non-Japanese speakers that assesses lexical knowledge of 10,000 Japanese words (The Japan Foundation and Japan Educational Exchanges and Services, 2012). The researcher assumed that these highest level words for non-native speakers of Japanese were still easy to recognize for Japanese native speakers and thus this baseline ensures that the present participants had a certain level of literacy as native-language processors.

In composing the English questions, the researcher paid special attention to the following issues:

(a) The length of definitions influences RTs so it must be equalized as much as possible. Therefore, all definitions consisted of two, three or four words.

(b) For all option words, the frequency band they originally belonged to in the Vocabulary Levels Test was redefined based on the JACET8000 word list. This word list reflects environmental traits of English education of Japan (JACET Committee of Basic Words Revision, 2003) therefore its frequency definition is assumed to be more appropriate for Japanese learners of English.

(c) Word items rated at the 6,000-word level or above in the JACET8000 were excluded in light of the findings by Aizawa (2006). He tested breadth of English vocabulary of Japanese college students using the JACET8000 and found that the average scores decreased as word frequency decreases but this trend stopped between the 4,000- and the 5,000-word levels. He suggested that the boundary of core vocabulary of this word list is around the 5,000-word level.

(d) Of course, the four word choices in a question should be of the same word frequency level. In addition, the researcher designed the word definition to consist of words in the designated frequency level or below. For instance, a 3,000-word level question consisted of four 3,000-level words as option words and a definition made up of words not exceeding this level. In this way, the difficulty of the designated word level is held constant.

A subtest for a particular frequency level consisted of 10 questions with the proportion of two-adjective, three-verb, and five-noun questions. This ratio followed the Vocabulary Levels Test.

The Japanese subtest followed the 2:3:5 ratio rule as well.

The questions were presented in a randomized order regardless of the distinction between English and Japanese and their frequency levels. The combination of a definition and the option words was fixed in the same question, but the arrangement of the option words also changed from time to time by generating random numbers. The random placement of option words was to minimize the placement effect on RTs; the first option word, for example, may be focused on earlier than the other option words, therefore a particular question could be answered faster, and vice versa. The testing items are listed in Appendix B.

The reading test. The reading test consisted of two sets of readings followed by comprehension questions taken from *Asian and Pacific Speed Readings for ESL Learners* (Quinn, Nation, & Millett, 2007). The original material included 20 sets of a 550-word readings with 10 multiple-choice comprehension questions. The passages were written with the most frequent 1,000 English words (p. 3). Because the present study focused on speed of reading, it was expected that, by using relatively easy readings, comprehension variability would be minimized and only speed variability would become salient.

The researcher examined the readability of all twenty readings in the original book by using Coh-Metrix version 2.0[1] (McNamara, Louwerse, Cai, & Graesser, 2005a, 2005b) and picked the two easiest level readings: Chapter-9 (“Jayaprana”) and Chapter-10 (“Buddhism”). (Flesch Reading Ease Score: Ch. 9 = 81.653, Ch. 10 = 81.775). The passages and the comprehension questions in the two reading sets were used without making any changes from the original.

The test started with displaying the “start” button in the middle of the screen. By clicking the button with a mouse, reading passages were presented and the internal stopwatch started up. The test taker could read over the whole passages by mouse operations. By clicking the “finish” button placed below the passages, the stopwatch halted and the reading time was recorded and then the screen changed into the comprehension question. The comprehension check was not timed. By pressing the “finish” button placed below the final comprehension question, the current reading set was completed and another reading set begun. The presentation order of the two reading sets was counterbalanced. Taking notes was not allowed throughout the reading session.

Data Collection

Data collection took place in a quiet classroom during a spring vacation. The researcher made contact with each participant via e-mail to decide the date and time for their participation. Each experimental session involved only one test-taker. The vocabulary test was given first, then the reading test. There were no breaks throughout the whole session. At the beginning of the vocabulary test, the test-taker was told to place the response pad on his or her dominant hand side of the computer. Also the researcher instructed participants to respond as fast as possible to the vocabulary questions. However, the fact that RTs were to be measured was not mentioned. Still, the researcher informed that some vocabulary questions might be presented in Japanese. This was to avoid an unwanted reaction lag when encountering an unanticipated type of question.

For the reading test, the researcher did not instruct the test-taker to read the passages as fast as possible but only notified them that some comprehension questions would be given. Their reading time would be faster if they were told that the reading was timed but the present study emphasized that comprehending the written sentences is an important nature of reading, in which reading with comprehension being neglected is not what we do in general. The researcher intended to measure reading time at the pace participants set in order to reach a certain degree of comprehension.

Data Screening Conditions

The study established a too-fast and a too-slow condition on an individual participant basis and his or her samples between the extremes were analyzed. The RT value for the too-fast condition was taken from the test-taker's fastest correct response in Japanese vocabulary questions. Since all participants are native speakers of Japanese, their fastest possible response must be seen in Japanese questions. Thereby, responses in English questions that were faster than the fastest Japanese response were discarded. Another extreme, the too-slow condition, was determined by calculating 3SD[2] based on his or her correct answers in English vocabulary questions.

Results

For all analyses reported below, the alpha level for significance was set at .05 and was adjusted by Bonferroni correction for the *t* tests that were performed separately based on the same variables. The effect sizes are reported as Cohen's *d* for the *t* tests and as eta squared (η^2) for the analysis of variance.

Effect of the Data Screening

The data screening procedure resulted in omitting a total of 14 correct responses from 705 correct responses of the vocabulary breadth test made by the 24 participants. About 2% of the all correct answers were excluded from analysis as unreliable. They included 10 too-fast responses and four too-slow responses.

Descriptive Statistics and Reliability Indices

Table 1 displays descriptive statistics of the vocabulary test and the reading test. The maximum possible was 40 for the vocabulary test and 20 for the reading test. The figures for the vocabulary test showed that no participant attained full marks; the ablest participant scored 38 out of a possible 40. The median score was 30.5, indicating that more than half of the participants achieved 75% accuracy. The median was slightly greater than the mean, indicating that the distribution was somewhat positively skewed. Still, all participants failed to answer some of the questions and in this sense, the test overall was moderately difficult for the participants. The figures for the reading test on the other hand showed most of the participants had understood the passages well: the median of the comprehension score was 19 out of a possible 20. The Kolmogorov-Smirnov test showed that the scores on the vocabulary test were normally distributed (p = .079, ns) and the reading comprehension scores were not normally distributed (p

< .001).

Table 1. *General results of the vocabulary test and the reading test*

	Vocabulary Test	Reading Test
Number of test-takers	24	24
Mean score	28.79	17.87
Median score	30.50	19.00
S.D.	7.75	3.02
Minimum score	4	8
Maximum score	38	20
Skewness	-1.74	-2.17
Kurtosis	3.79	4.39

Reliability indices of the vocabulary test and the reading test were calculated by means of the split-half comparison and the Cronbach's alpha reliability index. The Spearman-Brown coefficients of the split-half comparison were .865 for the vocabulary test and .812 for the reading test. Cronbach's alpha indices were .901 and .838, respectively. These figures indicated that the vocabulary test and the reading test in the present study were substantially reliable.

RQ1: Variability of RTs by Vocabulary Size

First, a correlation analysis was performed for all participants (N = 24) whose estimated English vocabulary size ranged from 500 to 4,800 words, and from 2.87 to 13.05 in the mean RT on the vocabulary test. Results showed that their vocabulary size and RTs were significantly and negatively correlated (r = -.613, p < .001; two-tailed). Overall, therefore, participants with a larger size of vocabulary chose correct choices more quickly than those with a smaller vocabulary size (Figure 1).

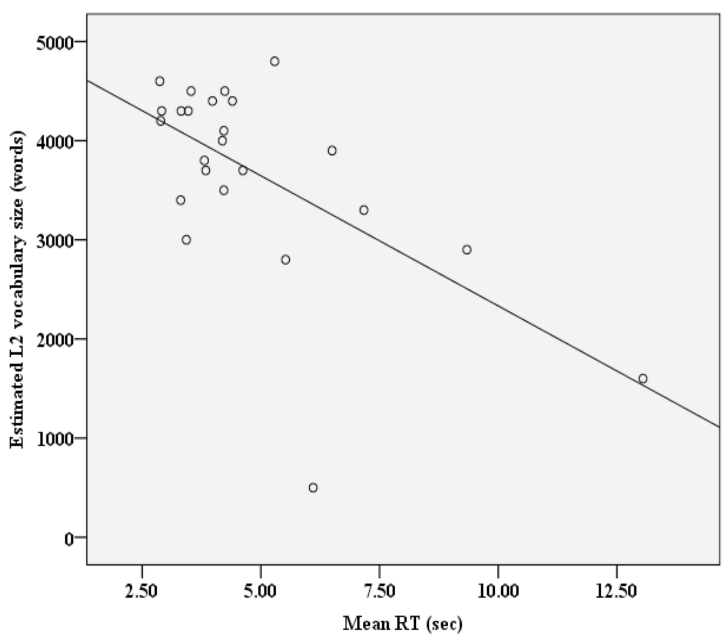


Figure 1. Correlation between vocabulary size and mean RT on the vocabulary test.

Next, the participants were divided into two groups, a group with a larger vocabulary size and a group with a smaller vocabulary size, by means of the median-split of their estimated L2 vocabulary sizes. Both groups had the same number of participants (*N* = 12). The mean scores of the two groups indicated that that the group with the larger vocabulary size had about 1,400 form-meaning connections more than the group with the smaller vocabulary size (Table 2)[3]; the group with the larger vocabulary size had about 4,441 words, ranging from 4,000 to 4,800, and the group with the smaller vocabulary size had about 3,075 words, ranging from 500 to 3,900. A *t* test revealed that the RT difference between the two groups was significant, $t(22) = -2.48$, $d = 1.01$, $p = .021$ (Bonferroni adjusted alpha = .025). Further, this difference was statistically clear in terms of the large effect size. The results overall showed that the number of words known clearly facilitated lexical accessibility in L2. Therefore the hypothesis for the first research question, whether learners with a larger vocabulary size answer faster on an L2 vocabulary breadth test than learners with a smaller vocabulary size, was supported.

Table 2. *Average number of words known and time of response*

	Vocabulary Test			
Group	Vocabulary Size		RT	
	Mean(words)	*SD*	Mean(sec)	*SD*
Higher half (*N*=12)	4,441	209.99	3.77	0.74
Lower half (*N*=12)	3,075	1,015.81	5.90	2.87

RQ2: Variability of RTs by Word Frequency

Figure 2 and Figure 3 illustrate the participants' overall performance at the 2,000-, 3,000-, 4,000- and 5,000-word levels on the vocabulary test. Based on these scores, the participants knew more words of higher frequency bands and their accuracy of vocabulary knowledge decreased as word frequency decreases, showing a decline from more to less frequent words (Figure 2). In contrast, there were no such linear trends on their mean RTs across the word frequency bands (Figure 3). As a result, the second hypothesis, that there are systematic increases in RTs as word frequency decreases in an L2 vocabulary breadth test, was not confirmed.

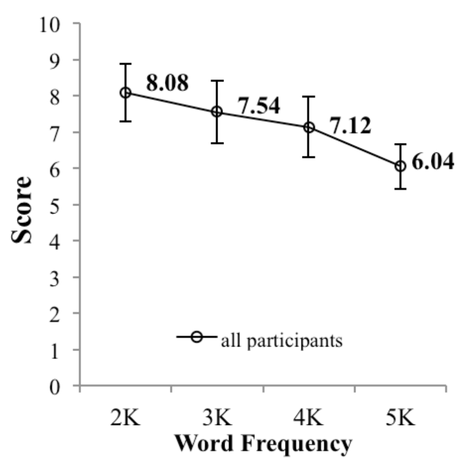


Figure 2. Mean scores on the vocabulary test by word frequency bands (all participants). Error bars indicate confidence intervals.

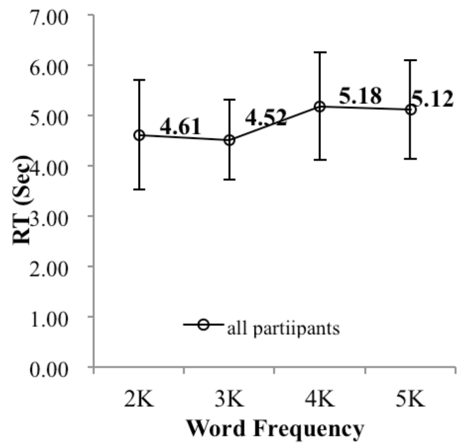


Figure 3. Mean RTs on the vocabulary test by word frequency bands (all participants). Error bars indicate confidence intervals.

In order to look into RT trends between different learners with different vocabulary sizes, the analysis proceeded by comparing the group with the larger vocabulary size and the one with the smaller vocabulary size, using a median-split in their vocabulary sizes, the same grouping employed in the analysis for RQ1. In addition to RT, score trends across the word frequency bands for both groups are also reported for reference. Note that one participant was excluded from the group with the smaller vocabulary size because no correct answers were made on the 3,000-word level questions.

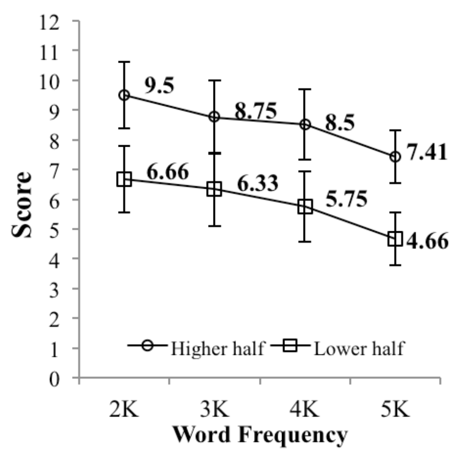


Figure 4. Mean scores on the vocabulary test by word frequency bands (grouped by vocabulary size). Error bars indicate confidence intervals.

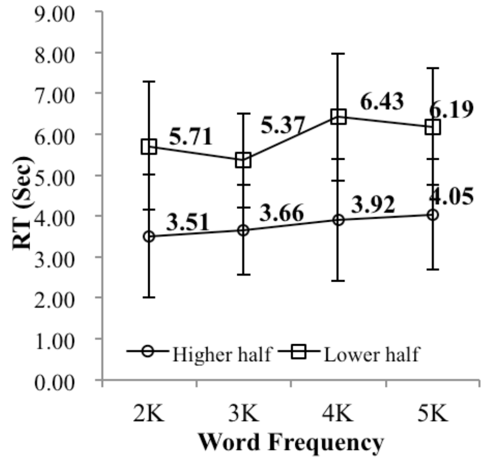


Figure 5. Mean RTs on the vocabulary test by word frequency bands (grouped by vocabulary size). Error bars indicate confidence intervals.

Overall, the mean vocabulary scores once again decreased as word frequency decreased. As illustrated in Figure 4, the two groups shared this tendency. However, they showed different trends on vocabulary RT (Figure 5). The group with the larger vocabulary size showed a gradual increase of RTs in accordance with word frequency but this was not the case for the group with the smaller vocabulary size. The reversals of RT in the group with the smaller vocabulary size was observed at two boundaries, between the 2,000- and 3,000-word levels (5.71 vs. 5.37), and the 4,000- and 5,000-word levels (6.43 vs. 6.19).

The RT differences were examined by a two-way factorial analysis of variance between Group (non-paired: larger vocabulary vs. smaller vocabulary) and Frequency (paired: 2K, 3K, 4K, and 5K) with the Bonferroni adjustment. Results showed a significant main effect for Group, $F(1, 21) = 5.59$, $\eta^2 = .192$, $p = .028$, but not for Frequency or the interaction between them. A post hoc comparison for the main effect on Group revealed that the statistical significance was due to the differences at the 3,000-word level (3.66 vs. 5.37; $F(1,21) = 4.968$, $p = .037$), the 4,000-word level (3.92 vs. 6.43; $F(1,22) = 5.101$, $p = .034$), and the 5,000-word level (4.05 vs. 6.19; $F(1,22) = 6.356$, $p = .019$).

As a result, a systematic gradual increase of RTs by decreasing word frequency was not confirmed statistically. However, it may be possible to say that a trend in this direction was observed for the group with the larger vocabulary size. The results collectively implied that the correlation between speed of semantic access to L2 words and word frequency depends on the vocabulary size of the learner. Therefore, the hypothesis for the second research question was partially confirmed.

Supplemental Analysis: Assessment of Lexical Access Automaticity

In order to examine whether the advanced participants' performance involved automaticity of lexical access, the coefficient of variation of RTs (CV_{RT}) was analyzed based on correct responses. CV_{RT} is calculated by the standard deviation of RTs divided by the mean RT (Segalowitz & Segalowitz, 1993). Therefore, a small value in CV_{RT} indicates that the RTs were stable over the trials, and they were overall quick when the mean RT was small. Although automatized knowledge is claimed to have the properties such that the processing is fast, ballistic, effortless, or unconscious (Segalowitz, 2003), CV_{RT} is only a useful index to estimate automaticity formation in the sense that we can observe even-paced responses quantitatively.

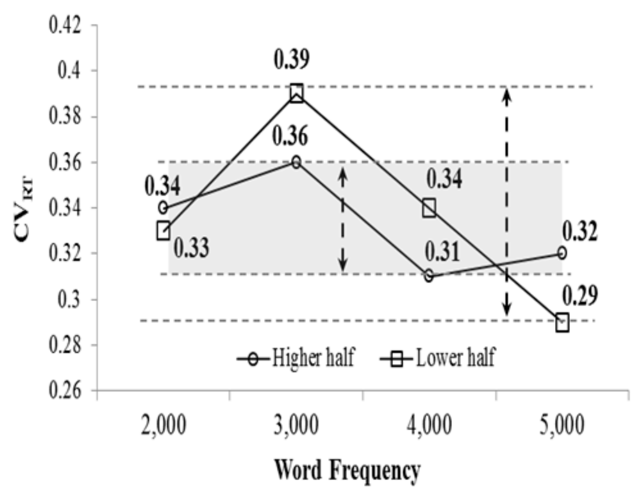


Figure 6. Coefficient of variation of RTs at each word frequency band.

Results showed that the mean CV_{RT} of the larger sized vocabulary group was smaller (CV_{RT} = .35) than that of the group with the smaller vocabulary size (CV_{RT} = .41). Overall though, the difference between them was not statistically significant, $t(22)=-1.47$, $p = .156$ (Bonferroni adjusted alpha = .025). As Figure 6 shows, correct responses of participants with a larger size of vocabulary were temporally less variable (from .31 to .36) than those of participants with a smaller size of vocabulary (from .29 to .39) across word frequency levels. This suggested that the larger sized vocabulary group had more automatic lexical access than the group with the smaller vocabulary size.

A small CV_{RT} value indicates automaticity in the sense that it reflects less variant RTs, but the present study also acknowledged that less variant CV_{RT} values across different word frequency levels indicate automaticity because speed of lexical access must be stable regardless of frequency of words in L1. However, none of the present participants had fully developed automatic L2 lexical access since their RTs and accuracy differed between word frequency levels, unlike in the L1, and in this sense all participants were at some lower developmental stages of automaticity. Still, the larger sized vocabulary group was considered to have superior automaticity of lexical access than the group with the smaller vocabulary size when their accuracy rate, smaller RTs, and stable CV_{RT} were collectively taken into consideration.

RQ3: The Relationship between RTs on a Vocabulary Test and Reading Speed

Prior to the analysis for the third research question, the researcher checked whether the present reading condition resulted in great variation in speed with high comprehension rate overall. As shown in Table 3, the two experimental groups showed at least 85% reading accuracy. Also, their mean comprehension scores were not significantly affected by their breadth of vocabulary

(18.83 vs. 16.91), *t*(22) = 1.60, *p* = .123, ns. Only words per minute in reading were significantly different between them, (150.16 vs. 106.04), *t*(22) = 3.54, *d* = 1.51, *p* = .002. As mentioned earlier, they read two sets of 550-word readings comprised of the most frequent 1,000 English words without time pressure. The average time spent on the reading test was 588 seconds (9 min. 48 sec.), and ranged from 353 seconds (5 min. 53 sec.) to 1,362 seconds (22 min. 42 sec.).

Table 3. *Average comprehension score and WPM in reading between groups*

	Reading Test			
Group Based on Vocabulary	Comprehension Score		Words per Minute	
Test Scores	Mean	*SD*	Mean	*SD*
Higher half (*N*=12)	18.83 (94.2%)	2.20	150.16	25.82
Lower half (*N*=12)	16.91 (84.6%)	3.50	106.04	34.49

The third research question was tested by means of a multiple regression analysis. In this analysis, participants who showed 90% accuracy or above in reading comprehension were analyzed (*N* = 20). The independent variables were the score and the RT of the vocabulary test. Because RTs of the vocabulary test were those of correct responses, the RTs represented both accuracy and speed of lexical-semantic access whereas the scores referred only to accuracy. The dependent variable was the words per minute of the reading test. Because participants to be analyzed were restricted to readers with high comprehension scores, as mentioned above, the variable must be a reflection of either accurate and fast reading or accurate but slow reading.

Table 4 shows the result of a multiple regression analysis for words per minute (WPM) in reading with the two explanatory variables, the vocabulary score and the vocabulary RT, provided by the forced entry method. The two predictor variables altogether explained 63.5% of the variance in the number of words read per minute. Further, the results indicated that the vocabulary RT alone could be a predictor of words per minute with a certain level of reading accuracy (*p* = .007) but the vocabulary score could not (*p* = .073, ns). However, a post hoc power analysis showed that the power of the analysis was .187, which is below the recommended level of 0.8, suggesting that the obtained results may have been based on data with a small sample size.

Table 4. *Multiple regression analysis for reading WPM*

Dependent variable: WPM in reading of above 90% accurate readers				
Procedure	Independent variable	β	*t*	*p*
Forced entry	Vocabulary RT	-.552	-3.100	.007
Forced entry	Vocabulary Score	.340	1.910	.073

An additional correlational analysis showed that the two predictor variables both had a relatively strong and significant association with words read per minute (*r* = -.746, *p* < .001 for Vocabulary RT vs. WPM; *r* = .654, *p* = .002 for Vocabulary score vs. WPM). The two predictor variables, the latencies in accurate responses and the rate of accuracy on the vocabulary test, were also interrelated significantly (*r* = -.569, *p* = .009). This conforms to our prediction that lexical knowledge is more accessible according as the number of words known increases.

As a result, it was suggested that speed of lexical semantic retrieval measured on a vocabulary test is predictive of values in determining accurate and fast reading comprehension. Therefore the hypothesis for the third research question, whether RTs of correct responses on an L2 vocabulary breadth test predict accurate and quick L2 reading, was supported.

Discussion

Vocabulary Size, Word Frequency and Speed of Semantic Access of Words

Our data obtained by the first and second research questions were overall in line with those demonstrated by previous studies. As a between-learner criterion, vocabulary size distinguishes speed of lexical access and as a within-learner criterion, higher frequency words are more readily accessed.

The present two groups compared were at the 4,000- and the 3,000-word level. Significant between-learner RT differences were found at the 3,000-word level and over. This suggests that the words in the 3,000-word level were still being acquired for the 3,000-word level learners and thus their lexical automaticity for those words had not been fully developed. Still it is viewed that, when compared with the 4,000-word level learners, the 3,000-word level learners were still in development of automaticity of lexical semantic access for the most frequent 2,000 words since the between-group RT difference on the 2,000-word level indicated near significance (Figure 5; 3.51 vs. 5.71; $F(1,22) = 3.955$, $p = .059$). This supports Laufer and Nation's (2001) finding: "increase in automaticity lags behind increase in size" (p. 23).

Our data also support Miralpeix and Meara's (2014) argument that automaticity of lexical access is observable for learners having a relatively large lexicon. The group with a large vocabulary size showed a stable transition of CV_{RT}s across the word frequency levels in the vocabulary test. Also, their overall RTs were significantly smaller than in the group with a smaller vocabulary size. However, it is argued that the advanced learners' automaticity was still developing because, for one thing, their RTs increased as word frequency decreases.

Characteristics of Multiple-Choice Formats as an RT Measure

The mean scores and RTs by word frequency in Figure 4 and Figure 5 showed a different tendency from those elicited on the yes-no format in previous research and this suggests unique characteristics of the multiple-choice format. Taking Harrington's (2006) results for example, scores and RTs showed basically the same tendency in which scores decrease and RTs increase as word frequency decreases (pp. 157–158). However, in the present study, only the group with large vocabulary size showed this effect on the multiple-choice format whereas the frequency effect on score was confirmed regardless of vocabulary size. This indicates that less proficient learners knew correct choices but for some reason they took more time to answer on the multiple-choice format.

The similarity of the trend in the scores between the yes-no format and multiple-choice formats is in line with previous findings. For example, Mochida and Harrington (2006) compared scores of a yes-no test with those of the Vocabulary Levels Test by using the same testing items. The yes-no test scores included raw scores and four kinds of adjusted scores by major scoring formulae. Results showed that all the five kinds of yes-no scores significantly correlated with the baseline scores measured by the Vocabulary Levels Test with correlation coefficients around $r = .80$. However, it should be noted that scores reflect the results of decision-making but RTs reflect the process of decision-making. Therefore, it is possible to think that multiple-choice formats impose more load on the test-takers than the yes-no format.

In relation to reaction time on vocabulary tests, Pellicer-Sánchez, and Schmitt (2012) attempted a unique scoring methodology for the yes-no format. Their assumption was that the test-takers must take time to answer when they were not sure about the answer. Therefore it may be possible to mark the exam based solely on their RTs if a certain RT threshold to ensure confident decisions could be determined. A methodological issue with the yes-no format is how to treat "yes" answers for non-words, often called false alarms. Ideally, the answer for non-words should be "no". Therefore even if learners said "yes" for non-words, such responses must take time and thus be captured by the RT methodology as unreliable answers. The presupposition here is that reliable test-takers must say "no" if they took a certain time to answer, whether it was for a non-word or not. However, the researchers found that the scores by the RT methodology were less accurate than those based on traditional formula-based scoring for non-native speakers, indicating that the learners said "yes" for non-words without taking time, unlike their original assumption. Therefore, the observed discrepancy of the RT trends of the low proficient learners is considered to be due to a difference in the difficulty of reaching a decision-making threshold using the two formats.

The present vocabulary test employed a definition-word association format therefore the minimum requirements to pick the correct choices are the knowledge of the correct option words and the knowledge of the words used in the definition sentences. Knowledge of distractors is not essential. In the case of the question in Figure 7, for example, the test-taker must know the meaning of *agreement* in the description and the meaning of the correct choice *contract*, and then have to judge that their meanings are proximate. However, it may be difficult for the test-taker to have enough confidence about the decision made on such least-required information because more appropriate words may be in the rest of the choices. Less proficient learners might face with this kind of situation. That is, learners were more confident in decision making on the multiple-choice format, whereas less confident intuitive decision making is more likely on the yes-no format. In this sense, the multiple-choice format most clearly identifies small sized lexicons.

Written agreement

1 factor 2 victim
3 region 4 contract

Figure 7. A sample question from the present vocabulary test.

Speed of Lexical Access and Speed of Reading

The multiple regression analysis for reading performance by the score and RT variables of the vocabulary test demonstrated that vocabulary scores alone do not predict speed of reading, and accurate and quick reading performance is associated with accurate and quick responses on a vocabulary test. Therefore, it is suggested that fast-accessible lexical knowledge facilitates reading performance. The present study also suggests that vocabulary breadth is an important factor for reading performance not only because comprehension is enhanced but also because lexical automaticity is developed as a function of increasing vocabulary and this is reflected in accurate and quick reading.

Educational implications from the findings of the present study are, first, developing breadth of vocabulary knowledge is important to lexical automaticity. Second, encouraging quick responses in a vocabulary test can be a part of reading training. For example, teachers of second language reading classrooms can include supplemental activities to expand their students' L2 vocabulary knowledge or tasks to train lexical automaticity as a part of training to develop fundamental skills in their L2 reading. Word recognition is one of the major lower-level processes in reading (Nassaji, 2014). Timed vocabulary tests, in this sense, are conceivable options for implementing these ideas in practical situations relatively easily. Finally, a use of timed lexical tests for placement purposes may allow us to make more precise classification of students. Since access speed is considered to be the upper criterion of word knowledge beyond the know-don't-know distinction, timed tests could be one of the options, especially for the placement of advanced learners.

Limitations

At least three major limitations need to be mentioned. First, the present study assumed that speed of lexical semantic access of L2 words could be measured by examining RTs on a vocabulary test and this could be an indicator of the degree of established semantic knowledge. On the other hand, the study also acknowledged that RTs do not illustrate the exact picture of lexical accessibility. To minimize this issue, a "pass" button was added on the response pad to avoid guessed answers and the researcher instructed test-takers to press the button when they did not know the answer. However, they did not follow this instruction at all because all participants made mistakes. Actually, only three of them used this button at least once. It could be interpreted that those participants had associated the word forms with the wrong meanings and thus they thought they knew the words since L2 learners' form-meaning links are often vaguely estabilished[4], but rather I suspect that the test-taking situation prompted them to overperform and this resulted in guessing. As a result, the "pass" button did not work in the way the researcher had intended.

Second, the estimated vocabulary sizes measured in the present study did not illustrate the test-takers' total breadth of L2 vocabulary. The vocabulary test consisted of the most frequent 5,000 words for the sake of measuring accessibility of commonly accommodated words for all learners. However it must be true that some participants knew more than that level. Therefore, their vocabulary sizes were underestimated in terms of their actual knowledge.

Lastly, the differences in vocabulary size of the current participants were not large, and therefore the same results may not be found with participants with different vocabulary sizes. Also, it should be acknowledged that the results do not fully illustrate the whole complex picture of lexical and reading competency since the present study focused on only a small part of the multilayered human language processing ability in a laboratory setting.

Conclusion

The present study proposed that *speed* is a valuable index of L2 lexical knowledge and showed that speed of accurate responses on a vocabulary test is a reliable indicator of fast, successful reading comprehension. Our goal in L2 learning is to make L2 knowledge practically available. Therefore, it is natural to think that a certain degree of speed is necessary. From this viewpoint, traditional paper-based vocabulary tests only measure learners' word knowledge with the know-don't-know distinction. By operationalizing temporal measures, as in the present study, vocabulary tests can classify L2 lexical knowledge in greater detail.

Acknowledgments

I would like to thank the anonymous reviewers and the editors for their valuable comments on earlier versions of this manuscript.

Notes

1. As of September 2014, *Coh-metrix 3.0* is available at http://cohmetrix.com/.

2. The formula for 3SD here is: Mean RT + (SD * 3).

3. The conversion formula from raw scores into vocabulary sizes is:
vocabulary size = ([2Kscore / 10] * 2000) + ([3Kscore / 10] * 1000) + ([4Kscore / 10] * 1000) + ([5Kscore / 10] * 1000)

4. For example, the Vocabulary Knowledge Scale (Wesche & Paribakht, 1996) assesses this possibility. This test requires the learner's self-rating about the degree of form-meaning knowledge of the target word based on a five-point scale, and there is an option "I have seen this word before, and I *think* it means _____" among them.

References

Aizawa, K. (2006). Rethinking frequency markers for English-Japanese dictionaries. In S. Ishikawa, K. Minamide, M. Murata, & Y. Tono (Eds.), *English lexicography in Japan* (pp. 108–119). Tokyo: Taishukan Publishing Company.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. A. Hutson (Ed.), *Advances in reading/language*

research: A research annual (Vol. 2) (pp. 231–256). Greenwich, CT: JAI Press.
Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing*, *16*, 131–162. doi:10.1177/026553229901600202
Currie, M., & Chiramanee, T. (2010). The effect of the multiple-choice item format on the measurement of knowledge of language structure. *Language Testing*, *29*, 471–491. doi:10.1177/0265532209356790
Educational Testing Service. (2015). *CEFR mapping study.* Retrieved August 4, 2015 from https://www.ets.org/toefl_itp/research
Fukkink, R. G., Hulstijin, J., & Simis, A. (2005). Does training in second-language word recognition skills affect reading comprehension?: An experimental study. *The Modern Language Journal*, *89*, 54–75. doi:10.1111/j.0026-7902.2005.00265.x
Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. *EUROSLA Yearbook*, *6*, 147–168.
Hulstijn, J. H., van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, *30*, 555–582. doi:10.1017/S0142716409990014
JACET Committee of Basic Words Revision. (Ed.). (2003). *JACET list of 8000 basic words*. Tokyo: The Japan Association of College English Teachers.
Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77. doi:10.1093/applin/21.1.47
Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.
Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook*, *1*, 7–28.
Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical test coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*, 15–30.
McNamara, D., Louwerse, M., Cai, Z. & Graesser, A. (2005a). *Coh-Metrix Version 1.4.* Retrieved December 22, 2012 from http://cohmetrix.memphis.edu
McNamara, D., Louwerse, M., Cai, Z. & Graesser, A. (2005b). *Coh-Metrix Version 2.0 Indices*. Retrieved December 22, 2012 from http://cohmetrix.memphis.edu
Miralpeix, I., & Meara, P. (2014). Knowledge of the written word. In J. Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 30–44). UK: Palgrave Macmillan.
Mochida, A., & Harrington, M. (2006) The yes/no tests as a measure of receptive vocabulary knowledge. *Language Testing*, *23*, 73–98. doi:10.1191/0265532206lt321oa
Moinzadeh, A., & Moslehpour, R. (2012). Depth and breadth of vocabulary knowledge: Which really matters in reading comprehension of Iranian EFL learners? *Journal of Language Teaching and Research*, *3*, 1015–1026. doi:10.4304/jltr.3.5.1015-1026
Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching*, *47*, 1–37. doi:10.1017/S0261444813000396
Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Boston: Heinle & Heinle.
Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Obunsha. (2011). *Japanese-Language Proficiency Test target 2,000: N1 vocabulary*. Tokyo: Obunsha Publishing.

Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring yes-no vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, *29*, 489–509. doi:10.1177/0265532212438053

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, *56*, 282–307. doi:10.3138/cmlr.56.2.282

Quinn, E., Nation, I. S. P., & Millett, S. (2007). Asian and Pacific speed readings for ESL learners: Twenty passages written at the one thousand word level. *English Language Institute Occasional Publication* (No. 24). Retrieved June 16, 2012 from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Speed-reading-whole.pdf

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, *19*, 12–25. doi:10.1177/003368828801900202

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge: Cambridge University Press.

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 Publishing.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). New York: Oxford University Press.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*, 369–385. doi:10.1017/S0142716400010845.

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, *19*, 53–67. doi:10.1017/S0142716400010572.

The Japan Foundation and Japan Educational Exchanges and Services. (2012). A summary of linguistic competence required for each level [PDF document]. Retrieved May 3, 2015 from https://www.jlpt.jp/e/about/pdf/comparison01.pdf.

van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Sims, A., Snellings., P. & Stevenson, M. (2004). Linguistic, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, *96*, 19–30. doi:10.1037/0022-0663.96.1.19

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, *53*, 13–40.

Yalin, S., & Wei, T. (2011). The relative significance of vocabulary breadth and syntactic knowledge in the predication of reading comprehension test performance. *Chinese Journal of Applied Linguistics*, *34*, 113–126.

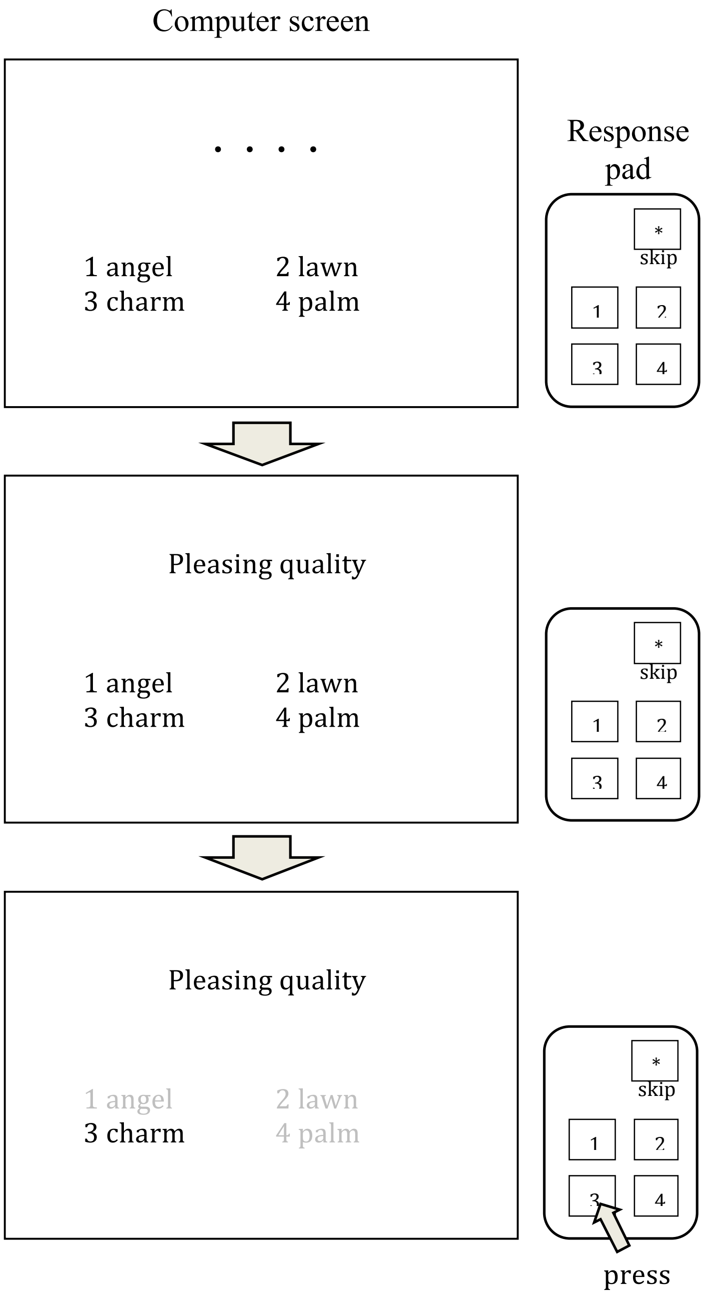
Yamashita, J. (2013). Word recognition subcomponents and passage level reading in a foreign language. *Reading in a Foreign Language*, *25*, 52–71.

Yildirim, K., Yildiz, M., & Ates, S. (2011). Is vocabulary a strong variable predicting reading comprehension and does the prediction degree of vocabulary vary according to text types. *Educational Sciences: Theory & Practice*, *11*, 1541–1547.

Appendix A

Presentation Procedure of the Vocabulary Test

1
2
3



1. ➢ First, a set of option words appears but the definition is hidden by dots. (Same words are presented in the same question but their placements (1 - 4) are randomized.)

2. ➢ Five seconds after, the definition is presented. At the same time, an internal timer count is started.

3. ➢ When an option word is selected by pressing the corresponding button, the other options are grayed out. Also the timer is stopped and response time is recorded but not displayed on the screen.

Appendix B

The Full List of the Vocabulary Questions

The 2,000-word Level

type	definition	distractor 1	distractor 2	distractor 3	correct choice
adj	wanting food	unique	specific	potential	hungry
adj	happening once a year	mental	magic	royal	annual
n	office worker	access	pen	media	clerk
n	written agreement	factor	victim	region	contract
n	person who is studying	journey	fund	gift	pupil
n	something you must pay	copy	layer	philosophy	debt
n	subject of discussion	evidence	schedule	contest	topic
v	break open	connect	urge	link	burst
v	make wider or longer	satisfy	grant	complain	stretch
v	be on every side	attach	convince	threaten	surround

The 3,000-word Level

type	definition	distractor 1	distractor 2	distractor 3	correct choice
adj	having no fear	mutual	brilliant	equivalent	brave
adj	best or most important	desperate	abstract	neutral	supreme
n	top of a mountain	empire	dust	horror	summit
n	stage of development	thread	file	fortune	phase
n	male or female	pond	salary	concrete	gender
n	study of the mind	motor	timber	charity	psychology
n	pleasing quality	angel	lawn	palm	charm
v	hurt seriously	grasp	wander	devote	injure
v	close completely	embrace	bond	demonstrate	seal
v	become like water	hire	illustrate	plead	melt

The 4,000-word Level

type	definition	distractor 1	distractor 2	distractor 3	correct choice
adj	most important	gross	explicit	controversial	primary
adj	last or most important	empirical	naked	objective	ultimate
n	good quality	exposure	analysis	mortgage	merit
n	chance of something happening	plot	draft	chill	prospect
n	part of a country	incidence	stability	integration	province
n	money paid for services	parameter	alcohol	vein	fee
n	agreement of permission	ward	institute	bishop	consent
v	fall down suddenly	launch	rape	assemble	collapse
v	expect something will happen	abolish	peer	oblige	anticipate

v	describe clearly and exactly	condemn	dispose	extract	specify

The 5,000-word Level

type	definition	distractor 1	distractor 2	distractor 3	correct choice
adj	next to	forthcoming	slim	profound	adjacent
adj	difficult to believe	striking	dim	tragic	incredible
n	loud deep sound	erosion	pulse	hydrogen	roar
n	a kind of stone	bulb	orientation	license	marble
n	circular shape	outfit	enforcement	scrap	loop
n	small stones	nail	ridge	exploitation	gravel
n	group of animals	mosaic	frost	stool	herd
v	say publicly	embarrass	coincide	shatter	proclaim
v	cut neatly	betray	penetrate	compile	trim
v	bring back to health	toss	skip	inquire	revive

First Language (Japanese)

type	definition	distractor 1	distractor 2	distractor 3	correct choice
adj	柔軟であるさま	あくどい	興味深い	生臭い	しなやか
adj	消えてなくなりやすい	ふさわしい	ややこしい	おろそか	はかない
n	非常にまばらな状態	ふきん	目の当り	怪談	過疎
n	付け加えること	申し出	味わい	訴訟	補足
n	物事が起ころうとする気配	意地	ボイコット	和洋折衷	兆し
n	果たすべきつとめ	車掌	博士	主導権	任務
n	人体に有害な物質	パトカー	職員	固体	毒
v	食事をととのえて出す	稼ぐ	見渡す	乗り出す	まかなう
v	仲立ちをする	転じる	痛む	訪れる	取り次ぐ
v	ねらいをはずす	押し付ける	やっつける	積む	そらす

The following 13 questions were used for the practice session and were not analyzed.

The 1,000-word Level

type	definition	distractor 1	distractor 2	distractor 3	correct choice
adj	not public	sorry	local	professional	private
adj	knowing what is happening	slow	independent	junior	aware
adj	liked by many people	firm	electric	previous	popular
n	not having something	area	accident	source	lack
n	part of 100	secret	parent	choice	percent
n	have a rest	operation	role	respect	relax
n	way of doing	trip	adult	attack	method

	something				
v	go up	introduce	design	concern	climb
v	make better	learn	limit	jump	improve
v	produce books and newspapers	release	elect	arrange	publish

First Language (Japanese)

type	definition	distractor 1	distractor 2	distractor 3	correct choice
adj	時間が長く経過すること	汚らわしい	くさい	おごそか	久しい
n	民間放送で流れる広告	酪農	要因	過失	コマーシャル
v	光を当てる	欠く	納める	交わす	照らす

About the Author

Masayuki Tanabe is currently a candidate in the Ph.D program in Second Language Studies in the Department of English at Aoyama Gakuin University, Tokyo, Japan. His research interests include second language vocabulary acquisition, second language vocabulary testing and cognitive processing in learning. E-mail: msyk@le.to