Measuring Japanese Learners' Lexical Accuracy and Fluency Using A Lexical Decision Task

Tohru Matsuo

語彙性判断テストを使用した日本人学習者の語彙の正確さと流暢さの測定

松　　尾　　　　　徹

Abstract

The purpose of this study is to measure word recognition fluency and accuracy for Japanese learners of English with a visual lexical decision task. This study focuses on investigating how reaction time and accuracy of visual word recognition in the lexical decision task differ depending on the participants' proficiency levels (Japanese university students, Japanese English teachers, and native speakers of English) and on the word frequency level (1K, 2K, 3K, and 4K) of the stimulus words. The results of a one-way and two-way mixed analysis of variance (ANOVA) showed that reaction time and lexical decision task accuracy generally distinguish the participants' proficiency. Also, a general frequency effect for reaction time was found for all groups.

Keywords: a lexical decision task, reaction time, word recognition

（Received September 21, 2018）

抄　　　録

本研究の目的は語彙性判断テストを用いての語彙認識の流暢さと正確さを測定することである。特に語彙性判断テストで測定する被験者の語彙の反応速度と正確さが被験者の英語能力（大学生、英語教員、英語母語話者）と語彙の頻度レベル（1000 語、2000 語、3000 語、4000 語）によってどのように異なるのかを調査する。分散分析の結果から語彙の反応速度と正確さは被験者グループの英語能力の差を区別することが判明した。また、全体的に頻度による語彙の反応速度の差がすべての被験者グループ間で見られた。

キーワード：語彙性判断テスト、反応速度、語彙認識

（2018 年 9 月 21 日受理）

1. Introduction

This study investigates Japanese learners' lexical accuracy and fluency, which is specified as the automaticity of written word recognition. In both L1 and L2 research, written word recognition, which refers to the processes of extracting lexical information from graphic displays of words (Koda, 2004), is widely considered to be one of the most important processes contributing to skilled reading comprehension (Grabe, 2009; Perfetti, 1999, 2007; Perfetti, Landi, & Oakhill, 2005). In addition, fluent reading is not possible without the automatic recognition of a large number of words (Grabe, 2009).

The concept of automaticity or automatization is based on skill acquisition theories, such as Anderson's adaptive control of thought theory (Anderson, 1983; Anderson & Lebiere, 1998) and Logan's (1988) instance theory. Anderson's theory regards automatization as starting with the conscious, controlled processing of declarative knowledge (i.e., knowledge of facts and rules, such as the knowledge of letter features, and letter-sound correspondences in the case of word recognition). After large amounts of processing, learners develop rapid, attention-free processing, which consists largely of routines characterized by "chunks" of elementary operations and computations (Anderson & Lebiere, 1998, p. 5). Alternatively, Logan's instance theory (1988) views automaticity as the retrieval of information from memory: "performance is automatic when it is based on single-step direct-access retrieval of past solutions from memory" (p. 493). Performance is initially executed based on rules (e.g., in English, the letter string *ph* corresponds to the phone /f/.); however, each time the individual utilizes a particular rule, the outcome is stored as a higher order unit called an instance. Hence, automatized performance is not based on the usage of a rule but on the retrieval of previously encountered instances. Regardless of the differences in these two theories of skill acquisition, automatic processing is viewed as effortless and fast, and is reflected in a decrease in latencies when performing a task.

Automatized word recognition is important for successful reading comprehension because of working memory limitations (Perfetti, 1985). Text comprehension is highly demanding and requires a high degree of cognitive control because readers must process multiple levels of language simultaneously (e.g., orthography, lexis, and morpho-syntax). If individuals have not automatized a large number of words, too many attentional resources must be devoted to recognizing individual words and sufficient cognitive resources are unavailable for higher order comprehension tasks, such as integrating new information with older knowledge, making inferences, and developing a representation of the text as a whole.

In order to measure automaticity of word recognition, researchers frequently use a computerized visual lexical decision task in which test-takers decide whether strings of letters are real words or non-words as quickly as possible. The lexical decision task produces two measures, reaction time and reaction accuracy. In Second Language Acquisition (SLA) studies, these two measures have often been utilized independently. For example, lexical decision task accuracy has been utilized to assess the vocabulary size of second language learners (e.g., Eyckmans 2004; Mochida & Harrington, 2006), whereas reaction time has been employed to determine the degree to which L2 learners' have developed automatic word recognition (e.g., Segalowitz & Hulstijin, 2005; Segalowitz & Segalowitz, 1993). However, few researchers have combined the two measures to determine how measures of reaction accuracy and response time simultaneously distinguish L2 learners' lexical proficiency (e.g., Harrington, 2006; Kojima, 2010). In this study I adapted Harrington's design (2006) and examined how well visual lexical decision task accuracy and response time discriminate English proficiency levels and item difficulty on the basis of word frequency with L1 Japanese English as a Foreign Language (EFL) learners.

2. Harrington's study

Harrington (2006) validated two measures of a lexical decision task, response accuracy and response time, and examined how the two measures served to discriminate among between-group proficiency levels and within-group levels of word frequency on an English lexical decision task. His research questions were whether lexical decision task accuracy improves as group proficiency and word frequency level increase, whether reaction time and lexical decision task accuracy improve as group proficiency and word frequency level increase, and whether lexical decision response stability improves as group proficiency and word frequency level increase. Response stability was measured with coefficient of variance of response time, which is calculated as the mean standard deviation of reaction times divided by the mean (Segalowitz, & Segalowitz, 1993). Three groups participated in the study: 32 intermediate English as a Second Language (ESL) students, 36 advanced ESL students, and 42 native speakers of English, who served as a control group. The 150-item lexical decision task contained 90 real words and 60 pseudo-words. The 90 real words consisted of 18 items from each of Vocabulary Levels Test's (Schmitt, Schmitt & Clapham, 2001) four frequency bands (2K, 3K, 5K, and 10K) and from the Academic Word List (Coxhead, 2000). The results indicated that reaction times systematically decreased as the participants' proficiency and word frequency increased, and accuracy increased as learner proficiency increased. The intermediate ESL group was less accurate and slower than the advanced ESL group, which was less accurate and slower than the native speakers of English. Moreover, both accuracy and reaction time measures discriminated among all word frequency levels (2K, 3K, 5K, and 10K word frequency levels) in both the intermediate and advanced ESL groups. Response variability, as measured by the coefficient of variance, also decreased as performance improved. However, the change was subtler than those of lexical decision task accuracy and response time. Harrington concluded that lexical decision task accuracy and response times can discriminate between proficiency levels and word frequency; therefore, these two measures of lexical knowledge can serve as valuable tools for examining the development of L2 proficiency.

3. The gaps and purpose of the study

This study is designed to address a gap in Harrington's study, which concerns the proficiency level of the participants. The participants in Harrington's studies have been ESL students whose English proficiency is higher than the Japanese EFL students in this study. In contrast, the university undergraduate students in this study have lower English proficiency and it is unclear to what extent the findings of previous researchers apply to them. Hence, the purpose of this study is to investigate whether the same trends observed in previous studies can be obtained with lower proficiency Japanese English learners.

The following research hypotheses were formed.

1. Lexical decision task accuracy of Japanese learners will improve as group proficiency level increases.
2. Lexical decision task accuracy of Japanese learners will improve as word frequency increases.
3. Lexical decision task reaction times of Japanese learners will decrease as group proficiency level increases.
4. Lexical decision task reaction times of Japanese learners will decrease as word frequency increases.

4. The study

4. 1. Participants

Three groups of participants took part in this study. The first group was 20 Japanese university students (male = 15, female = 5) majoring in law, whose English proficiency ranged from pre-intermediate to intermediate according to the university's classification system. They were first and second year students who had received formal English education for 9 or 10 years, and their approximate age was 19. Their mean Test of English for International Communication Institutional Program (TOIEC IP) score was 495 (Range: 265-525; *SD* = 73.74). The second group was 20 high proficiency Japanese English high school or university teachers (male = 5, female = 15), who have earned an Master of Art (MA) in Teaching English to Speakers of Other Language (TESOL) or Applied Linguistics. The last group was 20 native speakers of English (male = 15, female = 5) teaching at Japanese universities. These participants all had an MA in TESOL and served as a reference group.

4. 2. Instruments

The lexical decision task was made up of 144 items. In order to avoid response bias (Jiang, 2012), which is a type of cognitive bias caused by asking for an unequal number of positive and negative responses, the number of words and pseudo-words was balanced (72 words and 72 pseudo-words). In Harrington's (2006) study, the 2K, 3K, 5K, and 10K frequency bands were utilized for real words. However, the Japanese university students in this study had lower English proficiency than his ESL participants; therefore, the 72 real words consisted of 18 items from each of the first four 1,000 word families (18 items × four word frequency bands = 72 total items) in the British National Corpus. Example items used in this study are *high* and *brief* (from the first 1,000 word band), *song* and *store* (from the second 1,000 word band), *wild* and *storm* (from the third 1,000 word band), and *crew* and *slope* (from the fourth 1,000 word band). Stimulus words in the first, second, and third 1,000 word families were selected from the Japanese students' familiarity rate list (Yokogawa, 2006). In the Japanese students' familiarity rate list, familiarity rate was defined as how often learners think they see or hear the words, not whether they know the meaning of the words. The list shows the familiarity rate for 3,000 words on a scale that ranges from 1.48 (*See or hear rarely*) to 6.92 (*See or hear frequently*). This list was developed through the following process. First, the most frequent 2,981 words were selected from the Kilgarriff Lemmatised Frequency List (Kilgariff, 1997), and 19 words (the days of the week and the names of the months) were added. The resulting 3,000 words were administered to 810 Japanese EFL students from ten universities in western Japan. They were asked to rate how frequently they think they see or hear the words using a Likert scale ranging from 1 (*Never see or hear*) to 7 (*Very frequently see or hear*).

Because the Japanese students' familiarity rate list (Yokogawa, 2006) does not include beyond the fourth 1,000 word families, the 18 stimulus words (6 adjectives, 6 noun, and 6 verbs) in the fourth 1,000 word families were selected from JACET 8,000 word list (Aizawa, Ishikawa, & Murata, 2005). The JACET 8,000 list contains 8,000 words considered to be important for Japanese learners to communicate in English. These 8,000 words were selected from the British National Corpus and sub-corpus, which contains 5.8 million words from the following genres: newspaper, TV program transcripts, Junior and Senior High School English textbooks, scientific articles, and literature for young people. These 8,000 words are divided into eight 1,000-word frequency levels. Level 1, the first 1,000 words, covers an average of 88.6% of the running words in a high school textbook, and the first eight 1,000 word bands cover 98.7% of the running words.

When the frequency of word family of the stimulus words is controlled, one common problem is that low frequency words are generally longer than high frequency words. Researchers have shown that longer words produce longer reaction times (RTs), whether length is defined in terms of the number of letters or the number of syllables (e.g., De Groot, Borgwaldt, & Van, 2002; Stenneken, Conrad, & Jacobs, 2007). In order to avoid confounding the effects of frequency and word length on reaction times, two lexical properties, the number of letters and syllables were controlled through choosing six sets of three stimulus words, each of which consisted of four, five, and six letters in each frequency level. Besides controlling the number of letters and syllables, the part of the speech of the word was also controlled. Among six sets of three stimulus words, two sets are adjectives, other two are nouns, and the rest are verbs so that each of the part of the speech includes the equal number. Pseudo-words were randomly generated with ARC Non-Word Database (Rastle, Harrignton, & Coltheart, 2002) controlling for number of letters and syllables and selecting orthographically existing onsets, bodies, and legal bigrams so that they approximately mirror real words (e.g., *luig*, *opie*, *reuth*, and *thafe*).

4. 3. Procedures

The Lexical decision task was conducted utilizing the computer program RT builder V4 in Lextutor (Cobb, 2012). The participants individually took the test on a laptop computer in a quiet room. They were told they would see the strings of letters that were either real words or non-words. These instructions were given in English to the native speakers of English and in Japanese to Japanese English teachers and Japanese university students. The participants were asked to judge as quickly and as accurately as possible whether they knew the word. After completing approximately 10 practice items with an oral explanation by the researcher, they began the test. On each trial, as soon as the participant pressed 2 on the keyboard a *wait* appeared on the screen. After a specified interval a word or pseudo-word randomly appeared on the screen. The participants responded *Yes* (i.e., It is a word in English) by pressing 1, and *No* (i.e., It is not a word in English) by pressing 3 on the keyboard. No feedback was given. Most participants completed the test in 7 to 10 minutes.

5. Results

The results are divided into two parts. In the first section, an overview of the results is provided and the outlying responses are discussed. In the second part, the hypotheses are examined.

5. 1. Primary Analysis

There was a Japanese university student whose accuracy rate was 68.06% (46 errors out of 144 items). This rate was considered to be fairly low. In a lexical decision task, participants with an error rate of 20 percent or higher are often excluded (Jiang, 2012). Moreover, the false alarm rate of this participant, which identifies pseudo-words as real words, was 55.5% (40 nonword items out of 72 were wrongly identified as real words). This rate was exceptionally high compared with the overall mean false alarm rate of 8% (approximately 5 items out of 72 non-words). Therefore, this participant was excluded from further data analysis. The remaining 59 participants' overall mean accuracy rate on the lexical decision task was 94.18% with accuracy rates ranging from 81.94% to 100%.

Test scores for correctly identified real words were calculated for each level (1K, 2K, 3K, and 4K) and for overall performance. In lexical decision experiments, outliers commonly occur due to attention lapses, false-starts, and anticipatory responses. Generally there are two widely used methods to define and identify outliers for visual lexical decision task (Jiang, 2012). The first is based on the absolute value, which defines cutoff points as any RT that is shorter than 300 milliseconds (msec), which is known as the low cutoff and longer than 2,500 msec (known as the high cutoff). None of the RTs in this study fell below the low cutoff or above the high cutoff point. An alternative method was to identify outliers is the use of standard deviation. Outliers are usually defined as any RT that is outside 2, 2.5, or 3 standard deviations of the mean RT of the same participant (Jiang, 2012). In order to avoid distorted results, outliers are customarily replaced (*e.g.*, Segalowitz & Segalowitz, 1993; Harrington, 2006). In this study, responses more than 2.5 SDs beyond individual mean RTs were replaced with the value at the 2.5 *SD* point. This affected less than 2% of the data across all the groups. Only correctly identified real words were included in the final reaction time analyses.

5. 2. Secondary Analysis

5. 2. 1. Hypothesis 1: Lexical decision task accuracy will improve as group proficiency level increases.

Table 1 shows the overall accuracy rate (both correctly identified real words and correctly rejected pseudo words) by group. As Table 1 shows, the native English speakers' accuracy rate was higher than that of Japanese English teachers, and the accuracy rate of the Japanese English teachers was higher than that of the Japanese university students. Cronbach's alpha reliability estimates were calculated for each Japanese L1 group: Japanese university students = .70 and Japanese English teachers = .82. Reliability estimates were not applicable to the native English speakers due to the lack of variance, that is, almost of all the native speakers responded accurately to the majority of the items.

Table 1. *Descriptive statistics of overall accuracy performance rates by group.*

Group	*N*	*M*	*SD*	*Min*	*Max*
Japanese University Students	19	89.59%	4.39	81.94%	96.53%
Japanese English Teachers	20	95.45%	3.83	84.03%	99.31%
Native Speakers of English	20	97.50%	1.54	94.44%	100.00%

A one-way analysis of variance (ANOVA) was conducted to examine differences among the groups' accuracy performance. Group differences were significant, $F(2,56) = 27.21$, $p < .001$. The strength of the relationship between overall performance accuracy and three groups, assessed by partial η^2, was very strong, with grouping factor accounting for 49% of variance of the dependent variable.

Follow-up tests were conducted to evaluate pairwise differences among the means. The variances among three groups ranged from 1.54 to 4.39, and the homogeneity of variance test was significant, $p < .001$. Therefore, Dunnett's C was used for the post-hoc comparisons. There were significance differences between the Japanese university students and Japanese English teachers, and between Japanese university students and native speakers of English, $p < .05$. However, the pairwise comparison between Japanese English teachers and native speakers of English was not significant. These results partially supported hypothesis 1: lexical decision task accuracy improved as proficiency level increased, but the difference between Japanese English teachers and native speakers of English was not statistically significant.

5. 2. 2. Hypothesis 2: Lexical decision task accuracy will improve as word frequency increases.

Table 2 shows the proportion of mean accuracy rate (*i.e.*, 1 is perfect) and standard deviation for correctly identified real words by frequency and group. As for Japanese university students and Japanese English teachers, the accuracy means discriminate between the two

Table 2. ***Accuracy rate for correctly identified real words by frequency level and group.***

		Accuracy rate	
		M	*SD*
1K	Japanese University Students	.98	.03
	Japanese English Teachers	.99	.03
	Native Speakers of English	.99	.02
2K	Japanese University Students	.95	.04
	Japanese English Teachers	.99	.02
	Native Speakers of English	.99	.03
3K	Japanese University Students	.92	.09
	Japanese English Teachers	.98	.04
	Native Speakers of English	.99	.03
4K	Japanese University Students	.87	.08
	Japanese English Teachers	.97	.06
	Native Speakers of English	.99	.03
Overall	Japanese University Students	.93	.05
	Japanese English Teachers	.98	.01
	Native Speakers of English	.99	.00

groups and across the word frequency levels, although the high standard deviation of the Japanese university students implies considerable individual variation in their responses. On the contrary, the accuracy scores of the native speakers group did not change across all frequency bands, which clearly indicates a ceiling effect.

In Figure 1, the Japanese university students group clearly shows the frequency effects on accuracy for correctly responding to real words because their accuracy decreased as the word frequency level decreased. The same tendency can be observed for the Japanese English teachers, as they moved from the 2K to the 3K to the 4K word frequency levels; however, the decrease in accuracy was smaller than that of the Japanese university students.

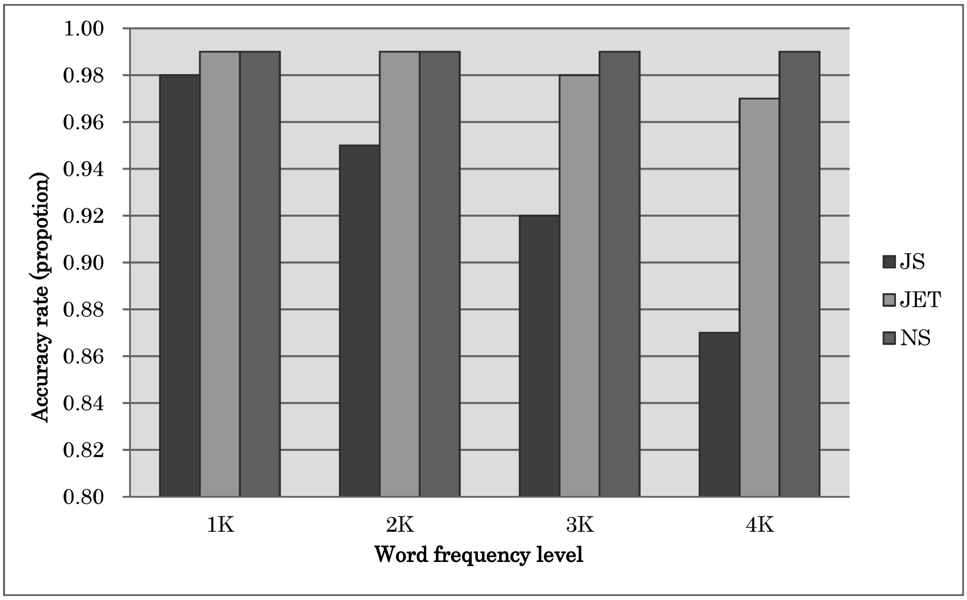

Figure 1. Mean accuracy scores by group proficiency level and word frequency level

The accuracy scores were analyzed using a mixed ANOVA for subjects and items. Group was the between-subjects factor (Japanese university students × Japanese English teachers × native speakers) and frequency level was the repeated measure factor (1K × 2K × 3K × 4K). The sphericity assumption, which hypothesizes that the variances of the data taken from the same participant are equal, was violated; hence, the results were reported using the Greenhouse-Geiser correction. Frequency effects on accuracy measure were significant, $F = 14.74$ (3, 168), $p < .001$, partial $\eta^2 = .21$. Tests of within-subjects contrast showed a linear relationship. Moreover, there were significant interactions between word frequency and group.

As a post-hoc analysis, all the pairwise comparisons for mean accuracy scores by frequency were conducted separately in each Japanese learner group. For the Japanese university students, F = 12.64 (3, 54), p < .001, partial η^2 = .41, significant differences were found for the 1K-3K, 1K-4K, and 2K-4K comparisons (p < .05, Bonferoni adjusted for multiple comparisons). However, even though the raw data showed a decrease in accuracy by word frequency levels for the Japanese English teacher, the effects were not significant, p = .061, partial η^2 = .143. Therefore, hypothesis 2 was supported only for the Japanese university students.

5. 2. 3. Hypothesis 3: Reaction times in the lexical decision task will decrease as group proficiency level increases.

Table 3 shows the means and standard deviations for the reaction times for correct responses to real words by group and frequency. As the overall column in Table 3 shows, the mean RT of Native English Speakers was faster than that of Japanese English Teachers, and the Japanese English Teachers were faster than Japanese University Students. A one-way ANOVA was conducted to examine group differences of overall reaction time to correctly respond to real words. The ANOVA was significant, $F(2,56)$ = 11.73, p < .001. Moreover, the strength of the relationship between overall reaction time and the three groups, assessed by partial η^2 was strong, as the grouping factor accounted for 30% of the variance of the dependent variable. Hence, hypothesis 3 was fully supported.

Table 3. ***Means and standard deviations for reaction time by group and frequency levels.***

		Reaction time (msec)	
		M	*SD*
1K	Japanese University Students	744	98
	Japanese English Teachers	679	82
	Native Speakers of English	650	92
2K	Japanese University Students	787	115
	Japanese English Teachers	742	99
	Native Speakers of English	636	81
3K	Japanese University Students	888	164
	Japanese English Teachers	777	101
	Native Speakers of English	681	99
4K	Japanese University Students	884	120
	Japanese English Teachers	826	166
	Native Speakers of English	702	99
Over all	Japanese University Students	825	113
	Japanese English Teachers	757	105
	Native Speakers of English	667	88

5. 2. 4. Hypothesis 4: Reaction times in the lexical decision task will decrease as word frequency increases.

As Table 3 shows, reaction times for the Japanese university students ranged from 744 msec at the 1K level to 884 msec at the 4K level. For Japanese English teachers, mean RTs increased from 679 msec at the 1K level to 826 at the 4K level. The Native Speakers of English had a mean RT of 650 msec at 1K and 702 msec at 4K. The scores were analyzed in a one-way mixed ANOVA for subjects and items. Group was the between-subjects factor (Japanese university students × Japanese English teachers × native speakers of English) and frequency levels was the repeated measure factor (1K × 2K × 3K × 4K). Frequency effects were significant, $F\,(3,54) = 51.44$, $p < .001$. Moreover, tests of within-subject contrasts were significant ($p < .001$), indicating the linear relationship among the four frequency bands. See Figure 2 for reaction time by group and word frequency.

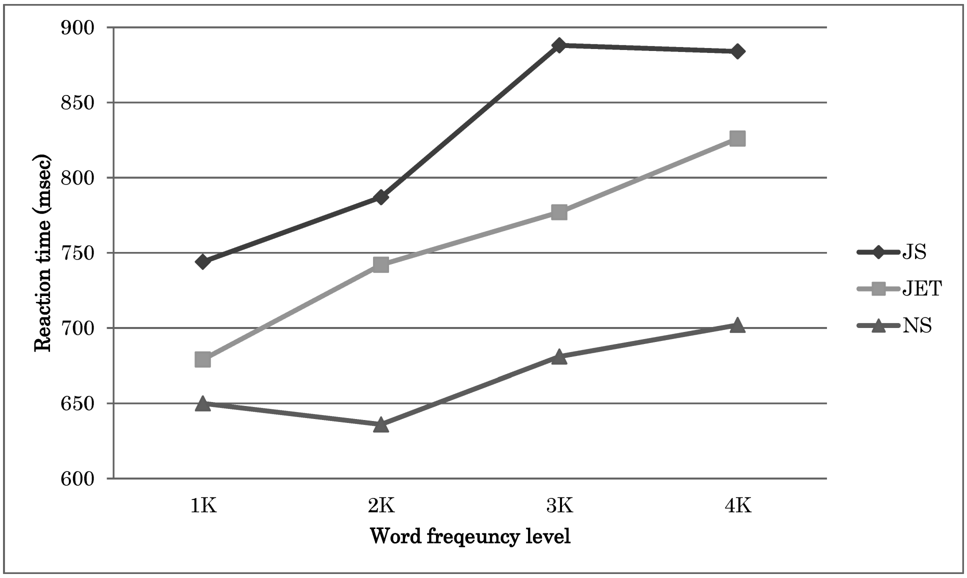

Figure 2. Mean reaction time by group proficiency level and word frequency level

In order to investigate the differences in reaction time among the four frequency bands, pairwise *t*-tests were conducted separately for each group. For the Japanese university students group, the results were significant, $F = 21.66\ (3, 54)$, $p < .001$, partial $\eta^2 = .55$. Significant differences were observed in all the pairwise comparisons except 1K-2K and 3K-4K. As for Japanese English teachers group, the results were also significant, $F = 28.83\ (3, 57)$, $p < .001$, partial $\eta^2 = .60$. Frequency effects on reaction time were observed in all the pairs except for 2K-3K and 3K-4K. For the native English speakers, $F = 15.52\ (3, 57)$, $p < .001$, partial $\eta^2 = .45$. Significant differences were observed at 1K-4K, 2K-3K, and 2K-4K. Table 4 shows the pairwise comparisons for reaction time by word frequency levels and group proficiency levels. These results generally supported hypothesis 4.

Table 4. ***Pairwise comparisons for reaction time by frequency and groups.***

Level	Reaction Time		
differences	*JS*	*JET*	*NS*
1K-2K	ns	*	ns
1K-3K	*	*	ns
1K-4K	*	*	*
2K-3K	*	ns	*
2K-4K	*	*	*
3K-4K	ns	ns	ns

Note
* = Difference significant at <.05, Bonferoni adjusted for multiple comparisons.
ns = Difference not significant. *JS* = Japanese university students.
JET = Japanese English teachers, *NS* = Native speakers of English.

6. Discussion

Overall accuracy (correct responses to real words and correctly rejected pseudo-words) generally improved as the participants' proficiency increased. Moreover, the standard deviation of the accuracy scores decreased as group proficiency increased. This indicated that the Japanese university students responded less consistently than the Japanese English teachers, who in turn were less consistent than the native English speakers.

Frequency effects on accuracy measures were not statistically significant for neither the Japanese English teachers nor the native English speakers. Native English speakers had approximately the same accuracy rate across the four frequency bands, a finding that indicated that their lexical knowledge was fully developed at least up to 4K. However, for Japanese English teachers, frequency effects on accuracy might become significant with a large sample size as the *p*-value was approaching significance ($p = .06$). On the other hand, frequency effects were observed for the Japanese university students at the 3K and 4K levels. This suggests that their lexical knowledge at 1K and 2K levels was fairly well developed; however, their lexical knowledge was not yet fully developed at the 3K or 4K levels.

Overall reaction time when correctly identified real words was significantly different among the three groups, a finding that clearly showed proficiency effects on reaction time. Japanese English teachers responded to real words faster than Japanese university students, and Native English speakers responded faster than the Japanese English teachers. This indicates that reaction time can perhaps discriminate the lexical proficiency of English learners, which supported the previous study (Harrington, 2006).

Frequency effects on mean reaction time were generally observed in all three groups, as all groups tended to respond more quickly to high frequency words than to low frequency words. The lack of a significant difference in mean reaction times at the 3K-4K levels for the Japanese university students group suggests that the 3K and 4K words were equally unfamiliar to them. For the Japanese English teachers, the insignificant value at 2K-3K would possibly become significant with more participants because the value was very close to significance ($p = .06$).

7. Conclusion

This study measured Japanese L1 EFL learners' accuracy response and reaction times, specifically to examine group proficiency and word frequency effect (up to the 4K levels) on accuracy and RT performance. Overall the results were aligned with the previous study (Harrignton, 2006), which showed that reaction times and lexical decision task accuracy generally distinguished proficiency. Moreover, a general frequency effect for reaction time was found for all groups. Even though this study was conducted with a relatively small sample size, RT research does not require a large number of participants scores (Jiang, 2012). There have been many RT studies in which no more than 20 participants were tested (*e.g.*, Costa & Santesteban 2004; Meunier & Segui, 1999). Jiang (2012) argued that in the case of single-presentation-list study, adding 10 participants did not change the pattern of the results from the ten people already tested except for increasing the power of the design. Moreover, Jiang further argued that a robust effect such as frequency effect takes a small number of participants to materialize. However, in future research, more participants with larger variance are needed in order to find clearer proficiency and frequency effects on accuracy and reaction time.

References

Anderson, J. R. (1983). *The architecture of cognition*. Mahwah, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah NJ: Erlbaum.

Aizawa, K., & Ishikawa. S., & Murata, T. (2005). JACET 8000 words. Tokyo: Kiriharashoten.

Balota, D. A., & Chumbley, J. I., (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology*, Human Perception and Performance. *10*(3), 340-357.

Cobb, T. (2012). Complete Reaction Timer (Version 4) [Computer software]. Retrieved from www.lextutor.ca/rt/

Costa, A. & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491-511.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly 34*, 213-238

De Groot, A. M. B., Borgwaldt, S., Bos, M., & Van den Ejinden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, *47*(1), 91-124.

Eyckmans, J. (2004). Measuring Receptive Vocabulary Size. LOT, Utrecht.

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

Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. In: Foster-Cohen, S., Kranjnovic, M.M., Digunovic J.M. (Eds.), EUROSLA Yearbook, pp. 147-168.

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

Kilgarriff, A. (1997). Putting frequencies in the dictionary. *International Journal of Lexicography*, *10*, 135-155.

Koda, K. (2004). *Insights into Second Language Reading*. New York, Cambridge University Press.

Kojima, M. (2010). Effects of word recognition speed, accuracy, and automaticity on reading ability. *Annual Review of English Language Education in Japan*, *21*, 151-160.

Kilgariff, A., 1997. Putting frequencies in the dictionary. *International Journal of Lexicography*, *10*(2), 135-155.

Logan, G. D. (1998). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492-527.

Meunier, F. & Segui, J. (1999) Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language*, *41*(3), 327-344.

Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing 26*, 73-98.

Perfetti, C. (1985). *Reading ability*. New York, NY: Oxford University Press.

Perfetti, C. (1999). Comprehending written language. A blueprint of the reader. In: C. Brown, P. Hagoort (Eds.), *Neurocognition of language*. Oxford: Oxford University Press, pp. 167-208.

Perfetti, C. (2007). Reading ability. Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357-383.

Perfetti, C., Landi, N., Oakhill,J., (2005). The acquisition of reading comprehension skill. In: M. Snowling., C. Hulme (Eds.), *The science of reading*. Malden, MA: Blackwell, pp. 227-247.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358, 534 nonwords: The ARC Nonword Database. Quarterly Journal of Experimental Psychology, 55A, 1339-1362.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*, 55-88.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In F. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179-201). Oxford: 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.

Stenneken, P., Conrad, M., & Jacobs, A.M. (2007). Processing of syllables in production and recognition tasks. *Journal of Psycholinguistic Research*, *36*(1), 65-78.

Yokogawa, H., (2006). 日本人英語学習者の英単語親密度文字編 [Familiarity rate of the English words for Japanese learners of written English word version], Tokyo: Kuroshioshuppan.