

Vocabulary Education & Research Bulletin

Summer 2019

Volume 8, Issue 1

This promises to be an exciting year in vocabulary, and VERB is likewise blessed with a wide variety of research. **Raymond Stubbe** and **Kousuke Nakashima** take a close look at active and passive recall test scores for Japanese students. **David Allen** investigates the proportion of English words with loanword equivalents in JHS textbooks. **Joseph Suzuki-Parker** and **George Higginbotham** examine whether method of administration affects word association test responses. **Brandon Kramer**, **Tohru Matsuo**, **Stuart McLean**, and **Steve Cornwell** look at targeted interventions for summer vacation vocabulary knowledge attrition. There are also two poster presentation synopses from the Vocabulary SIG Forum in November. **Tim Stoeckel** and **Dale Brown** introduce a test of extremely high-frequency words, and **Magda Kitano** and **Katsuhiko Chiba** attempt to encourage reluctant students to study for weekly vocabulary tests. Enjoy!

Magda Kitano & Tim Stoeckel, VERB editors

Table of Contents

Short Papers

| | |
|---|----|
| Raymond Stubbe & Kousuke Nakashima | 2 |
| <i>Examining the Difference between Active and Passive Recall Test Scores for Japanese Students when an L1 - L2 (Japanese-English) Bilingual Format is Used</i> | |
| David Allen | 7 |
| <i>Identifying Loanwords in a Junior High School Textbook Wordlist</i> | |
| Joseph Suzuki-Parker & George Higginbotham | 11 |
| <i>Does Method of Administration Influence Word Association Test Responses?</i> | |
| Brandon Kramer, Tohru Matsuo, Stuart McLean & Steve Cornwell | 17 |
| <i>Measuring Attrition of L2 Receptive Vocabulary Knowledge Over the Summer Vacation</i> | |

Poster Presentation Synopses

| | |
|--|----|
| Tim Stoeckel & Dale Brown | 23 |
| <i>Introducing a Test of the First 44 Members of the New JACET 8000</i> | |
| Magda L. Kitano & Katsuhiko Chiba | 29 |
| <i>Motivating Low-Achievement Learners to Prepare for Weekly Vocabulary Tests: The Come-Back Vocabulary Test</i> | |
| SIG News | 35 |

Short Paper**Examining the Difference between Active and Passive Recall Test Scores for Japanese Students when an L1 - L2 (Japanese-English) Bilingual Format is Used**

Raymond Stubbe raymondstubbe@gmail.com

Kousuke Nakashima nakacy8587@hotmail.com

It is generally accepted in the field of second language (L2) vocabulary acquisition that active recall of a word's form is significantly more difficult than passive recall of that word's meaning. According to the degrees of strength of knowledge hypothesis (Laufer & Goldstein, 2004, pp. 406-408), *recalling the form* of an L2 word based on a first language(L1) prompt (active recall) is a more challenging task than *recalling the meaning* (in L1) of a word based on an L2 prompt (passive recall). Laufer and Goldstein (2004), testing primarily Hebrew and Arabic learners of English, reported that their bilingual active recall (L1 to L2) vocabulary tests were significantly more difficult than the bilingual passive recall (L2 to L1) tests. This study reported 39.4% higher scores for the passive recall test over the active recall test, on average (see Table 1). This study gave tests to four different groups, with *N* sizes of 134, 106, 82 and 113, for a total of 435, an average *N* size of 108.75. Each group was tested on 30 words from different frequency levels: the 2K level (being the 1001-2000 most frequent words in English); the 3K level; the 5K level; and words from the Academic Word List (AWL; Coxhead, 2000), respectively. In this study, the first letter in the answer portion was provided for both test formats "to prevent the student from supplying nontarget words that have the same meaning" (Laufer & Goldstein, 2004, p. 406). An example of an active recall question (p. 430) for the item "Victory" was:

"V _____ (ניצחון)".

An example of a passive recall question (p. 431) was:

"Victory _____ נ".

Twelve years later, Laufer and McLean (2016), testing two distinct groups of English learners, were interested in the effect of loanwords on various testing formats including active and passive recall. This study, using monolingual (L2 - L2) recall tests,

reported 13.1% higher scores on average for the passive recall test over the active recall test, with their Hebrew students on the 80 tested items (see Table 1). This figure is much lower than the 39.4% average reported in Laufer and Goldstein (2004). The other group of tested learners were Japanese students, whose passive recall scores were only 1.2% higher on average than their active recall scores, likely a non-significant difference.

Table 1

Average Active Recall versus Passive Recall Means from Laufer and Goldstein (2004) and Laufer and McLean (2016)

| Study | N | k | AR mean | PR mean | diff PR - AR | AR | PR | % diff |
|--------|--------|----|------------|------------|--------------------|--------|--------|--------|
| | | | | | | mean % | mean % | |
| L&G | 108.75 | 30 | 4.39 | 16.22 | 11.83 | 14.6% | 54.1% | 39.4% |
| L&M -H | 79 | 80 | 25.15 | 35.59 | 10.44 | 31.4% | 44.5% | 13.1% |
| L&M -J | 58 | 80 | 23.78 | 24.76 | 0.98 | 29.7% | 31.0% | 1.2% |

Note. AR denotes Active Recall; PR denotes Passive Recall; diff denotes difference; % diff denotes diff / k; L & G denotes Laufer and Goldstein (2004), figures are averages of 4 tested frequency levels; L&M denotes Laufer and McLean (2016), -H denotes Hebrew speakers, -J denotes Japanese speakers, these figures are averages of the loanword and non-loanword test results.

Judging by the passive recall percent scores, the items tested in Laufer and Goldstein’s (2004) study appear to have been generally easier for the Hebrew test-takers than the items used in the Laufer and McLean study for the Hebrew test-takers (54.1% versus 44.5%). However, the active recall percent scores suggest the opposite (14.6% versus 31.4%). Whereas bilingual (L1 - L2) tests were used in Laufer and Goldstein (2004), both the active and passive recall tests used in Laufer and McLean (2016) were monolingual English (L2 - L2) tests. This may account for the higher active recall percent mean in the later study. For Laufer and McLean’s active recall test the first 3 (or fewer) letters of the target word together with its definition in simplified English (the L2) were provided. An example of an active recall question for the word “juvenile” was:

“They are still juv_____. (young, not adult)”.

An example of a passive recall question was:

“Someone who is juvenile is still” (Laufer & McLean, 2016, p. 205),

with students being asked to supply English equivalents.

This L2 to L2 format may account for a portion of the much smaller difference in

scores between the two test formats for the Hebrew students and especially the Japanese students, compared to Laufer and Goldstein (2004) (see Table 1, far right column). Having to supply an L2 answer in the passive recall test may have been considerably more difficult than providing an L1 translation would have been for these Japanese test-takers. It is also possible that providing up to two extra letters at the beginning of the target word in the active test may have activated the students' word knowledge more than the single first letter provided in Laufer and Goldstein (2004), thereby contributing to the smaller differences in active versus passive scores.

The aim of the current was to examine the difference between active and passive recall test scores for Japanese students when a Japanese-English (L1 - L2) bilingual format is used, to determine if the results are more in line with the Laufer and Goldstein (2004) findings than the monolingual (L2 - L2) format results reported in Laufer and McLean (2016).

Methodology

Two vocabulary tests were created: an L1 to L2 (Japanese to English) active recall test of 50 English words, and a passive recall (L2 to L1) test of the same 50 items. Items were selected from the students' textbook and randomly ordered in each test. For the active recall test questions, a Japanese translation was provided, followed by the second letter of the target word to avoid correct but nontarget answers. It was felt that supplying the second letter would provide less of a hint than supplying the initial letter. An example of such a question was:

弱い: _e_____.

The answer is "weak", and the blank line following the 'e' does not give an indication of the number of letters to follow. An example of a passive recall test questions was:

weak _____ ,

with two blanks provided to allow for multiple meanings.

Following Laufer and Goldstein (2004) and Laufer and McLean (2016), participants ($N = 54$ low-proficiency English learners) took the active recall test first, at the beginning of class. Towards the end of that same class, students took the passive recall test. Forms were collected immediately following each test. Between the tests, students participated in a 45-50 minute English lesson. The active recall tests were graded by one of the authors, using a liberal marking system. Spelling errors were considered acceptable

because we were interested in whether or not the students knew a correct meaning for the item. Thus, grades were awarded dichotomously (1 or 0) with no partial credit. The passive recall tests were similarly graded by a native Japanese teacher of English.

Results

Table 2 presents the descriptive test results for the two formats. The mean for the active recall test was 24.8, while the passive recall test mean was considerably higher at 33.7. A paired *t*-test found that that the difference between the means was statistically significant, $t(103) = 8.76$, $p < .001$, confirming the hierarchy of difficulty outlined in Laufer and Goldstein (2004). Finally, the percentage difference between the means (17.8%) is more in line with Laufer and Goldstein's (2004) statistically significant difference of 39.4% than Laufer and McLean's (2016) non-significant difference of 1.2%.

Table 2

Descriptive statistics for the active recall and passive recall tests.

| Test | Mean | SD | Range | Mean % |
|----------------|-------|------|---------|--------|
| Active recall | 24.80 | 6.17 | 13 - 37 | 49.6% |
| Passive recall | 33.69 | 6.18 | 18 - 44 | 67.4% |
| Difference | 8.89 | | | 17.8% |

Note: $k = 50$; $N = 54$.

Conclusion

Laufer and Goldstein (2004) found large differences between their bilingual active recall and passive recall test scores, in line with their degrees of strength of knowledge hypothesis. However, in Laufer and McLean (2016), the differences between their monolingual active recall and passive recall formats with the Japanese participants were very small. The present study set out to examine the difference between active and passive recall test scores for Japanese students when a Japanese-English (L1 - L2) bilingual format is used. Results suggest that when a bilingual testing format is employed, active recall test scores are significantly lower than passive recall scores, in line with the degrees of strength of knowledge hypothesis.

References

- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
<https://doi.org/10.2307/3587951>
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436. doi:10.1111/j.0023-

8333.2004.00260.x

Laufer, B., & McLean, S. (2016). Loanwords and vocabulary size test scores: A case of different estimates for different L1 learners. *Language Assessment Quarterly*, 13(3), 202-217. doi: 10.1080/15434303.2016.1210611

Short Paper**Identifying Loanwords in a Junior High School
Textbook Wordlist**

David Allen allen.david@ocha.ac.jp

Background

The literature on Japanese loanwords deriving from English (e.g., テーブル / teeburu / *table*) suggests that they are generally a beneficial resource for Japanese learners of English (Allen, 2018; Daulton, 2008). However, research has yet to show how this benefit can be maximized. The first step towards investigating the impact of loanwords in teaching and learning contexts is to identify how many of the English words found in learning materials have loanword equivalents in Japanese. Here, I address this issue regarding the vocabulary taught in English textbooks in junior high schools (JHS) in Japan.

Aims

In this paper, I utilize a wordlist derived from a corpus of JHS English curriculum textbooks published in Japan and Japanese corpus frequency data to answer the following question: What proportion of English words in JHS textbooks have loanword equivalents in Japanese?

Method

To investigate the proportion of English words that have loanword equivalents in JHS textbooks, I obtained the wordlist from the *Junior High School English Textbook Corpus* (JHSETC; Northbrook & Conklin, 2018a, 2018b). The corpus contains a total of 152,966 tokens and is made up of all of the junior high school English curriculum textbooks approved by the Ministry of Education, Culture, Sports and Sciences (MEXT) for the year 2013 (*Columbus 21 English Course, New Crown English Series, New Horizon English Course, One World English Course, Sunshine English Course, and Total English New Edition*).

The initial wordlist extracted from the corpus contained 6681 distinct word forms. I cleaned this list by removing the following: nonword strings arising from automatic text conversion (e.g., 1a2b), Japanese words, proper nouns, non-lexicalized utterances (e.g., *huh*), and abbreviations. All regularly inflected words were then combined into word families following Bauer and Nation's (1993) Level 2 classification

and the frequencies were summed (e.g., 66 occurrences of *table* and four occurrences of *tables* were combined into one entry, *table*, which had 70 occurrences). The resulting word list contained 2,843 entries.

To identify loanwords in the JHSETC wordlist, I used a wordlist derived from the *Balanced Corpus of Contemporary Japanese* (BCCWJ; National Institute for Japanese Language and Linguistics, 2013). The BCCWJ wordlist contains English ‘lemmas’ for all loanwords, which signify the English word from which the loanword derives (e.g., *table* for テーブル). Using the VLOOKUP formula in Excel, English words from the JHSETC were looked up in the BCCWJ wordlist, and their corresponding Japanese loanwords were identified along with their frequency data.

Table 1

English Word and Japanese Loanword Frequency Data for a Sample of 14 Words from the JHSETC

| Word | JHSETC frequency | BNC-COCA band | BCCWJ | | |
|-------------|------------------|---------------|----------|-----------|-----------------------------|
| | | | Loanword | Frequency | Frequency per million words |
| rhythm | 6 | 3k | リズム | 2214 | 21.2 |
| order | 16 | 1k | オーダー | 1066 | 10.2 |
| natural | 8 | * | ナチュラル | 596 | 5.7 |
| highway | 2 | 3k | ハイウエー | 364 | 3.5 |
| purple | 2 | 2k | パープル | 230 | 2.2 |
| balloon | 7 | 4k | バルーン | 134 | 1.3 |
| cheer | 9 | 2k | チアー | 73 | 0.7 |
| airport | 22 | * | エアポート | 36 | 0.3 |
| garbage | 9 | 6k | ガーベッジ | 14 | 0.1 |
| continent | 9 | 3k | コンティネント | 3 | 0.0 |
| tenderness | 1 | * | NA | 0 | 0 |
| grandmother | 53 | * | NA | 0 | 0 |
| belong | 3 | 2k | NA | 0 | 0 |
| hymn | 4 | 6k | NA | 0 | 0 |

Note. Asterisks denote words which do not have their own entries in the BNC-COCA lists because they are derivatives or compounds of separate headwords (e.g., *natural* is a derivative of *nature*).

Results

Of the 2,843 list entries, 2,090 (74%) had a loanword equivalent in the BCCWJ

corpus. Also, 1281 (45%) had a loanword equivalent occurring at or above once per million words. For illustration, Table 1 displays every 200th word from the list ordered in terms of Japanese cognate frequency. This table shows the English word and its frequency in the JHSETC corpus as well as the frequency level in the British National Corpus – Corpus of Contemporary American English word family list (BNC-COCA; Nation, 2012). The corresponding Japanese loanword and its frequency in the BCCWJ are displayed if the English word was identified in the BCCWJ corpus.

Conclusions

These data show that the majority (74%) of English words in the JHSETC wordlist have loanword equivalents and just under half (45%) have loanword equivalents that are relatively frequent in Japanese. These proportions in the JHSETC are almost identical to the proportions of English words that have loanwords in the most frequent 3000 words in the BNC-COCA corpus (i.e., 73% and 44%; Allen, 2018). Moreover, the proportion of English words in the JHSETC corpus with Japanese loanwords that occur at or above 1 per million words in the BCCWJ is almost the same as that reported in Daulton (2008), who showed that 45% of the most frequent 3000 words in the BNC had a loanword equivalent in Japanese. In sum, the words in JHS textbooks, which are mostly high frequency, have the same proportion of loanwords as that of the high-frequency words in the English language.

Future Directions

An important subsequent question is, which of these loanwords are likely to be known to JHS students? Based on familiarity ratings of loanwords and frequency data, Allen (2018) proposed that loanwords occurring at or above 1 per million words should be known in Japanese to most university undergraduates (this would correspond to 45% of the words in the JHSETC). However, whether this is also true for JHS students is unknown. Because JHS students are less experienced readers, they will have less loanword knowledge, which means a larger proportion of loanwords occurring above this threshold is likely to be unknown. Future research should therefore investigate the level of loanword knowledge of JHS students in relation to the BCCWJ frequency data.

Knowledge of different senses of loanwords, as well as JHS learners' ability to use loanword knowledge effectively to aid English communication, are also key areas for future research. In addition, future research should consider the degree of phonological and semantic similarity between the English and Japanese words because these variables may influence learners' understanding and use of the words in both languages (Allen, 2018; Allen & Conklin, 2014).

References

- Allen, D. (2018). The prevalence and frequency of Japanese-English cognates: Recommendations for future research in applied linguistics. *International Review of Applied Linguistics in Language Teaching*. Advance online publication. doi: 10.1515/iral-2017-0028
- Allen, D., & Conklin, K. (2014). Cross-linguistic similarity norms for Japanese-English translation equivalents. *Behavior Research Methods*, 46(2), 540-563.
- Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279.
- Daulton, F. E. (2008). *Japan's built-in lexicon of English-based loanwords*. Clevedon: Multilingual Matters.
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Unpublished paper. Retrieved from www.victoria.ac.nz/lals/about/staff/paul-nation
- National Institute for Japanese Language and Linguistics. (2013). *BCCWJ Word List*. Retrieved from https://pj.ninjal.ac.jp/corpus_center/bccwj/en/freq-list.html
- Northbrook, J., & Conklin, K. (2018a). Is what you put in what you get out? Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics*, Advance online publication. doi: 10.1093/applin/amy027
- Northbrook, J., & Conklin, K. (2018b). “What are you talking about?” *International Journal of Corpus Linguistics*, 23(3), 311-334.

Short Paper

Does Method of Administration Influence Word Association Test Responses?

Joseph Suzuki-Parker torin2582@icloud.com

George Higginbotham george@edu.hkg.ac.jp

Word association tests (WATs) have been employed for over 100 years, in various forms, due to their capacity to tap into the subconscious network of words that constitute an individual's mental lexicon. In this period, tests have been administered in different ways for different purposes – more recently as a way to measure the productive vocabulary knowledge of second language (L2) learners. There has been standardization of parts of the methodology and a gradual improvement in our understanding of how the different formats of the test impact responses. However, despite efforts to address methodological issues, such as how to select stimulus words (SWs) and how to classify responses (Fitzpatrick, 2006; Nissen & Henriksen, 2006), inconsistencies between studies remain. The current study builds on our understanding of how WATs work by exploring the effect of *method of administration* (MOA). Specifically, do learners respond in different ways to written and verbally administered WATs? With an MOA effect having been observed in studies of native speakers (Entwisle, 1966; Linton & Brotsky, 1969) it was hypothesized that an MOA effect would also be observed with L2 learners in modern WAT formats. This paper reports preliminary findings.

Word Association Tests

As one of the eight aspects of word knowledge identified by Richards (1976), word association is a key component in L2 language proficiency. The most direct way to access the relationships between words in the learner's mental lexicon is by administering a WAT. This usually involves a participant being shown (or hearing) a SW, and orally responding with (or writing down) the first word that comes to mind. By examining the relationship of responses to a list of SWs, researchers can make inferences about the structure and development of the participant's mental lexicon.

While WATs have long been used to inform first language (L1) studies, following the publication of an influential study by Meara (1983), there has been considerable progress in developing WATs for use with L2 learners. Meara outlined many of the methodological issues that needed to be addressed if the full potential of WATs were to be realized in an L2 context. Since then a number of studies have attempted to re-evaluate

basic assumptions and resolve Meara's issues in an effort to bring consensus to the field (Fitzpatrick, 2006, 2007, 2009; Nissen & Henriksen, 2006). A variable yet to be explored within L2 WAT research, however, is method of administration (MOA). Particularly, whether or not the kind of responses given during a WAT are affected by the test being administered in an oral or written format. If there is an effect, controlling for the way a WAT is administered will allow researchers to interpret data more accurately. The current study therefore compares the responses from (a) a format which employs a written stimulus and requires a written response (WW) and (b) one which provides an aural stimulus and requires an oral response (AO), to answer the question: Is there a qualitative difference between word association responses elicited using different methods of administration?

The Current Study

The participants were undergraduate students attending a private Japanese university. As a check on familiarity with the SWs, participants' receptive vocabulary knowledge was tested using a modified version of the New General Service List Test (NGSLT; Stoeckel & Bennett, 2015). Ten questions were selected from each of the first four levels of the test (up to the 2,240 word-frequency rank), for a total of 40 questions. In order to maximize the total number of known SWs, subjects who scored less than 50% on the NGSLT ($n = 3$) were excluded from the study, resulting in a final n -size of 15 (5 female, 10 male).

Of the potential test formats, the WW and AO methods were selected for the present study, as (based on a large L1 study by Linton & Brotsky, 1969) those formats were expected to differ the most. A total of 18 SW pairs (i.e. 18 SWs for each test) were selected from among the first 2,240 frequency-based entries in the New General Service List (NGSL; Browne, Culligan, & Phillips, 2013). Strict control over variables known to affect WAT responses was achieved by matching SW pairs on each test for word frequency, word class, word length and syllable count (Fitzpatrick, 2006, 2009; Higginbotham, 2014). In addition to the above criteria (following Wolter, 2001), common Japanese loan-words such as *aachisuto* (artist), and words likely to elicit stereotypical/unproductive responses (e.g. *cat*→*dog*) were avoided. The proportion of word classes selected within each test was as follows: nouns = 50%, verbs = 22%, adjectives = 16%, and prepositions = 11%. Except for prepositions, this closely resembles the proportions found in Zareva (2005).

For each method, a worksheet was prepared to elicit a single response for each SW, with SW pairs split across the two worksheets. Both tests were administered to all 15 participants by a single examiner over a two-week period. The first week, the written

test was administered as a group, and worksheets were completed by each participant. The following week, the oral test was administered individually with the interviewer completing the worksheet based on each participant's oral responses. After the completion of both tests, data were collected, and post-test interviews were conducted to help classify ambiguous responses. Finally, adopting the categories of word association outlined in Fitzpatrick (2006), responses were then classified as either *meaning*-based associations (super-ordinates, synonyms, etc.), *position*-based associations (collocates, compounds, etc.), *form*-based associations (words that share orthographic or phonological properties), or *other* (erratic or no response).

Results

General Results

Including erratic and no response types from the *other* category (64 responses), a total of 540 responses (WW = 270, AO = 270) from 15 participants to two WATs were collected and classified. Figure 1 presents the number of responses in each category for each WAT. A number of observations can be made. First, while the WW resulted in a predominance of *meaning*-based associations (145), the AO resulted in a predominance of form-based associations (89). Second, the WW outnumbered the AO in *meaning*-based associations (WW = 145, AO = 84); however, the AO outnumbered the WW for the remaining *position*-based (AO = 59, WW = 34), *form*-based (AO = 89, WW = 65) and *other* categories (AO = 38, WW = 26). In order to determine if the observed differences were statistically significant, the four categories were subjected to paired samples *t*-tests.

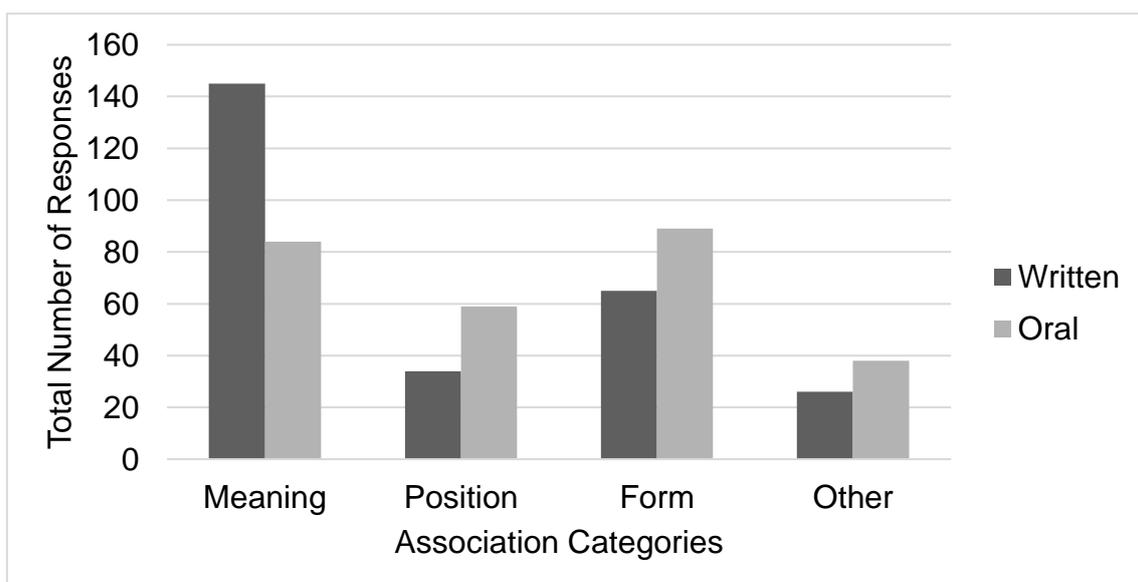


Figure 1. Number of responses for each association Category.

Quantitative Results

Table 1 presents the descriptive statistics of each category for both methods. In the meaning-based category there was a significant difference found between the WW ($M = 9.66$, $SD = 2.82$) and AO ($M = 5.60$, $SD = 2.16$) formats, $t(14) = 4.70$, $p < .001$, $d = 1.20$. In the *position*-based category a significant difference was found between the WW ($M = 2.26$, $SD = 1.94$) and the AO ($M = 3.39$, $SD = 2.15$) formats, $t(14) = 3.51$, $p = .003$, $d = .80$. The *form*-based category also yielded a significant difference between the WW ($M = 4.33$, $SD = 2.49$) and the AO ($M = 5.93$, $SD = 2.98$) formats, $t(14) = 2.20$, $p = .02$, $d = 0.50$. There was not a significant difference between the two formats in the *other* category. Based upon Sawilowsky's (2009) and Cohen's (1988) conventions, the effect size for the difference in mean scores for the *meaning*-based category was very large, that for the *position*-based category was large, and that for the *form*-based category was medium.

Table 1
Descriptive Statistics for WW and AO WAT ($n = 15$)

| Category | MOA | Min | Max | Mean | SD | SE |
|----------|-----|-----|-----|------|------|------|
| Meaning | WW | 5 | 14 | 9.66 | 2.82 | 0.73 |
| | AO | 2 | 11 | 5.60 | 2.16 | 0.56 |
| Position | WW | 0 | 7 | 2.26 | 1.94 | 0.50 |
| | AO | 1 | 8 | 3.93 | 2.15 | 0.56 |
| Form | WW | 0 | 9 | 4.33 | 2.49 | 0.66 |
| | AO | 1 | 11 | 5.93 | 2.98 | 0.77 |
| Other | WW | 0 | 3 | 1.73 | 1.10 | 0.27 |
| | AO | 0 | 6 | 2.53 | 1.96 | 0.50 |

Discussion

The current study set out to examine the possibility that word association responses of L2 learners could be affected by how a WAT is administered, with an eye to improving our understanding of WAT methodology. A comparison of mean values between the two methods revealed significant differences in the main association categories, with the largest effect size found between the responses in the (semantically related) meaning-based category. This suggests that the organization of links in the L2

mental lexicon could be partly dependent on the medium (written or oral) of language reception. This finding has implications for future WAT studies. If confirmed, researchers will need to consider effects related to MOA before comparing WAT research results, in order to avoid a situation of comparing ‘apples to oranges’. Additionally, differences found between the two methods could partially account for the variation in results reported in previous L2 WAT studies.

While these findings suggest that MOA does have an effect on responses, due to its small scale, the current study cannot be viewed as anything more than a preliminary investigation. To confirm these tentative findings, a similar study with a larger number of participants and more SWs is planned. The literature suggests that a suitable number of participants would be around 30 to 40 (Fitzpatrick, 2009; Higginbotham, 2014) and a suitable number of stimulus words would be between 40 (Higginbotham, 2014) and 100 (Fitzpatrick, 2009), depending on whether retrospective interviews are used. Numbers such as these will increase the statistical power of the study, permit an analysis of classification subcategories, and allow comparisons between subjects following Fitzpatrick’s (2007) individual profiling style of analysis.

References

- Browne, C., Culligan, B., & Phillips, J. (2013). *New General Service List*. Retrieved November 18, 2018, from <http://www.newgeneralservicelist.org/>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Entwisle, D. R. (1966). *Word associations of young children*. Baltimore: John Hopkins Press.
- Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. *EUROSLA Yearbook*, 6(1), 121–145.
- Fitzpatrick, T. (2007). Word association patterns: unpacking the assumptions. *International Journal of Applied Linguistics*, 17(3), 319–331.
- Fitzpatrick, T. (2009). Word association profiles in a first and second language: Puzzles and problems. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical Processing in Second Language Learners* (pp. 38–52). Bristol, Blue Ridge Summit: Multilingual Matters.
- Higginbotham, G. (2014). *Individual profiling of second language learners through word association*. (doctoral Dissertation) Retrieved from <https://cronfa.swan.ac.uk/Record/cronfa42500>.
- Linton, M. L., & Joyce Brotsky, S. (1969). Administrative procedures in the word association test: Method of stimulation and method of responding (oral vs.

- written). *Journal of Verbal Learning and Verbal Behavior*, 8(1), 123–128.
- Meara, P.M. (1983). Word associations in a foreign language. *Nottingham Linguistic Circular*, 11, 28–38.
- Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics*, 16(3), 389–408.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10, 77–89.
- Sawilowsky, S.S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 597–599.
- Stoeckel, T., & Bennett, P. (2015). A test of the New General Service List. *Vocabulary Learning and Instruction*, 4(1), 1–8.
- Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A Depth of individual word knowledge model. *Studies in Second Language Acquisition*, 23, 41–69.
- Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System*, 33(4), 547–562.

Short Paper**Measuring Attrition of L2 Receptive Vocabulary
Knowledge Over the Summer Vacation**

Brandon Kramer Brandon.L.Kramer@gmail.com

Tohru Matsuo tohruwombat@me.com

Stuart McLean stumc93@gmail.com

Steve Cornwell cornwell@wilmina.ac.jp

Background

Most students at Japanese universities attend classes for about eight months of the year, with breaks during the summer and early spring for around two months each. During this time students often take a break from studying and focus on working part-time jobs, relaxing with friends, or travelling. When removed from the school context rich in foreign language stimulus, students might be expected to lose access to some of the knowledge they have gained, in a process called attrition. The degree of lexical attrition during these breaks is of great concern, and any interventions found to impede this attrition would be valuable for all stakeholders.

Aims

While previous researchers have studied summer L2 lexical attrition (for summaries see Bardovi-Harlig & Stringer, 2010; Grendel, Weltens, & de Bot, 1993), very few have examined the effects of targeted interventions on preventing this loss. This study aims to address that gap by investigating the state of L2 receptive vocabulary knowledge before and after summer vacation. Specifically, we were interested in the following treatments: (a) digital paired-associate studying of the target words, which would promote access through spaced retrieval (Nakata, 2011); (b) extensive reading, because “literacy supports retention and impedes attrition” (Bardovi-Harlig & Stringer, 2010, p. 25); and (c) travelling abroad, which should increase exposure to the L2 target vocabulary.

Research Questions:

RQ1: Can attrition be found in Japanese university students’ L2 receptive vocabulary knowledge after a two-month summer break?

RQ2: Is there a relationship between changes in L2 receptive vocabulary knowledge and time spent using a digital paired-associate vocabulary study application, words read through extensive reading, or experience travelling abroad?

Methods

Context

The context for this study was a private women's university in western Japan with an institution-wide vocabulary program that encourages the deliberate studying of the most frequent words of English from the New General Service List (NGSL; Browne, Culligan, & Phillips, 2013). Using a freely available web application (Memrise.com), students studied the NGSL words in a paired-associate format utilizing spaced repetition (for more details on the benefits of such applications, see Nakata, 2011). Throughout the first semester, the students studied a continually expanding range of words (60 new words each week, for 13 weeks) and were tested with two productive review quizzes each week (one written and one spoken) within different classes of an integrated curriculum. At the end of the first semester, all first-year students had reviewed, in principle, the first 1,280 words of the NGSL (not including the first 500 words, which were only studied by the remedial classes). In addition to the vocabulary program, all first-year students were required to complete an extensive reading requirement (180,000 words were necessary for full points) using the web application Xreading.com, where they were exposed to similar words to those studied from the NGSL. After this first semester finished, the students were encouraged to review studied words with the Memrise app and read books from Xreading.com in their free time over the summer vacation.

Procedures

Three target vocabulary words were chosen randomly from each of the weekly ranges. Two rounds of tests measuring these target words ($K = 39$) were given to four intact classes of first-year Japanese university students, and only those participants who were present for both administrations were retained for analysis ($N = 103$; 18-19 years old; TOEIC range: 155-550). For each administration, a meaning-recall test was given first, followed by a four-choice meaning-recognition (i.e., multiple-choice) format (see example items in Figures 1 & 2). This order was chosen to avoid a test-effect where the participants could learn the meanings from the multiple-choice distractors, influencing the results of the meaning-recall test. The tests were administered digitally via SurveyMonkey.com, and the questions within each test format were randomized for each participant to reduce cheating. Participants were told they could skip questions they did not know the answer to, and each administration took about 30 minutes. The first administration was given between July 23rd and July 27th, 2018, and the second administration was given immediately upon the start of the second semester, September 24th to 28th.

2. solution: This is one solution.

Figure 1. Meaning-recall format used in this study.

14. solution: This is one solution.

要素

予算

解決法

提案

Figure 2. Meaning-recognition format used in this study.

Time spent studying vocabulary during the final month of summer vacation was found for each student through Memrise. Unfortunately, more detailed information than this was unavailable, because the application lacked a dedicated Learner Management System. Words read through the school's extensive reading program were measured using the word counts from Xreading. As per the system's settings, only the words for those books the student was able to demonstrate adequate understanding of were counted, meaning they were able to score at least 60% on the five-item comprehension check. Finally, travelling abroad information was self-reported using a questionnaire given following the final test administration in September. The students were asked to list any trips they took to a foreign country, and any which were English-speaking countries were dummy-coded for analysis.

Results

The meaning-recall data were scored dichotomously by a single native speaker of Japanese, also fluent in English. The items were scored according to whether the participants could demonstrate knowledge of the target vocabulary as used in the example sentences. For instance, even if the core meaning of the word was correct, it was not awarded a point when the part of speech was incorrect. An additional rater rated 10 randomly selected questions to conduct a rater reliability analysis ($\kappa = .618$; agreement = 81.0%). Descriptive statistics for all tests can be seen in Table 1.

For both the recall and recognition tests, a paired-samples *t*-test was conducted using JASP (JASP Team, 2018) to determine if there was a difference between pre- and post-test measures (see Table 2 and Figure 3). A significant difference ($p < .05$) was found between the meaning recognition tests, with the participants scoring slightly higher on the post-test ($M = 32.5$) than the pre-test ($M = 31.7$), indicating a gain in knowledge.

Table 1

Pre- and Post-Test Scores for Meaning Recall and Meaning Recognition Vocabulary Tests

| Vocabulary Measure | Min | Max | <i>M</i> | <i>SEM</i> | <i>SD</i> | Skewness | Kurtosis | α |
|-----------------------|-----|-----|----------|------------|-----------|----------|----------|----------|
| Recall Pre-Test | 4 | 35 | 18.8 | 0.73 | 7.36 | 0.10 | -0.67 | .877 |
| Recall Post-Test | 4 | 37 | 18.4 | 0.72 | 7.27 | 0.22 | -0.52 | .872 |
| Recognition Pre-Test | 20 | 39 | 31.7 | 0.46 | 4.72 | -0.60 | -0.33 | .811 |
| Recognition Post-Test | 19 | 39 | 32.5 | 0.44 | 4.52 | -0.99 | 0.63 | .816 |

Note. *N* = 103; *SES* (standard error of skewness) = 0.24 for all tests; *SEK*(standard error of kurtosis) = 0.47 for all tests; α = Cronbach's alpha.

Table 2

Paired Samples T-Tests between Pre- and Post- Measures

| Vocabulary Measure | <i>t</i> | <i>p</i> | <i>M</i> | <i>SE</i> | 95% <i>CI</i> | | <i>d</i> |
|----------------------|----------|----------|----------|-----------|---------------|-------|----------|
| | | | | | Low | High | |
| Recall Pre-Post | 0.94 | 0.350 | 0.38 | 0.403 | -0.42 | 1.18 | 0.09 |
| Recognition Pre-Post | -3.10 | 0.002 | -0.86 | 0.279 | -1.42 | -0.31 | -0.31 |

Note. Two-way Student's *t*-tests of pre- minus post-test scores, thus positive values for the mean difference (and effect size *d*) indicate the presence of attrition, or a decrease in test scores; *M* = mean difference; *SE* = standard error of the mean difference; *d* = Cohen's *d*.

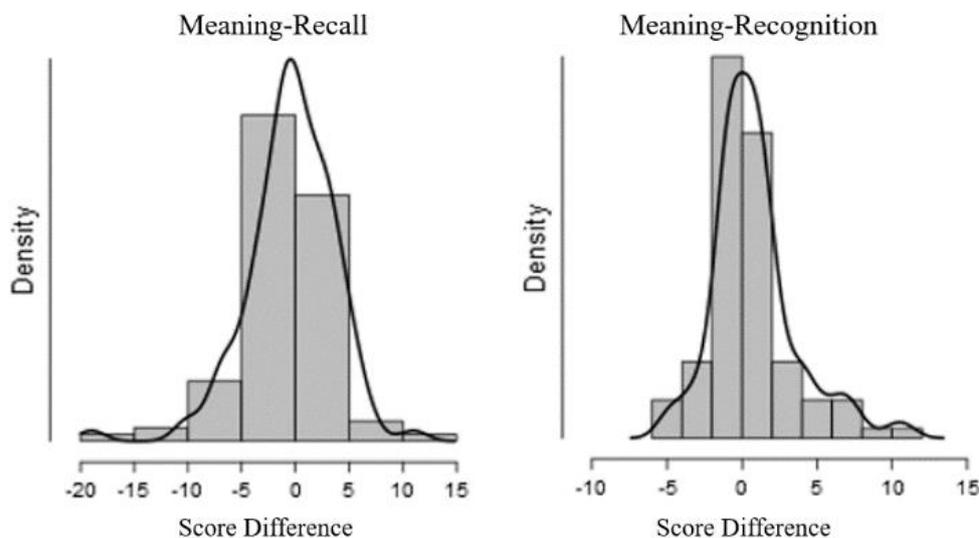


Figure 3. Distributions for the differences between pre- and post-test scores for both item formats.

Despite everyone being encouraged to do so, only some students studied with Memrise (*n* = 41) and Xreading (*n* = 60) over the summer break (see Table 3 and Figure

4 for descriptive statistics). Furthermore, only 11 students were classified as having travelled abroad to an English-speaking country. For each test format (recall and recognition), a multiple linear regression was calculated with JASP to predict the change in vocabulary scores based on minutes studied using Memrise, words read in Xreading, and whether the participant travelled abroad. The variables were added in order of predicted importance (Memrise first, followed by Xreading, and finally travelling abroad), but none were significant predictors of score changes, with only 2.3% of variance explained, $F(3, 98) = .78, p = .508$.

Table 3

Descriptive Statistics for Memrise and Xreading Usage Over the Summer Break

| Treatment | <i>n</i> | <i>Min</i> | <i>Max</i> | <i>M</i> | <i>SEM</i> | Median | <i>SD</i> | Skew | <i>SES</i> | Kurt | <i>SEK</i> |
|-----------|----------|------------|------------|----------|------------|--------|-----------|------|------------|-------|------------|
| Memrise | | | | | | | | | | | |
| Minutes | 41 | 2 | 12,740 | 494.5 | 311.9 | 40 | 1,997 | 6.06 | 0.37 | 37.82 | 0.72 |
| Xreading | | | | | | | | | | | |
| Words | 60 | 181 | 148,900 | 35,600 | 4,780 | 19,990 | 37,030 | 1.25 | 0.31 | 1.24 | 0.61 |

Note. *SES* = standard error of skewness; *SEK* = standard error of kurtosis.

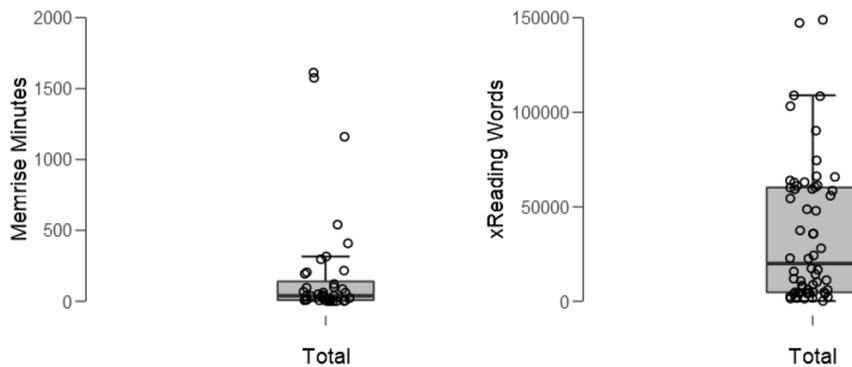


Figure 4. Boxplots showing Memrise and Xreading use over the summer break (one extreme outlier who practiced for 12,738 minutes was removed from the Memrise graph and the multiple linear regression analysis).

Conclusions

Investigating RQ1, we were unable to find evidence of attrition in receptive vocabulary knowledge after two months of summer vacation. Rather, a significant, albeit small ($d = -.31$; Plonsky & Oswald, 2014), gain in meaning recognition scores was found. Furthermore, no meaningful relationships were found between individual differences in

pre- and post-test receptive vocabulary scores and the number of hours spent studying vocabulary through Memrise, the number of words read through Xreading, or experience travelling abroad over the summer holiday period (RQ2). Overall, these results support the arguments of some researchers that receptive vocabulary knowledge is not easily lost relative to productive knowledge (Schmitt, 2010; Weltens & Grendel, 1993).

Future Directions

Future research in this area should utilize productive measures of vocabulary knowledge in addition to other, more sensitive, measures of receptive vocabulary knowledge such as reaction time experiments or tests of collocational knowledge (as advocated by Weltens & Grendel, 1993). Furthermore, ensuring greater compliance with the recommended summer studying would perhaps aid statistical analyses by providing greater variance.

References

- Bardovi-Harlig, K., & Stringer, D. (2010). Variables in second language attrition: Advancing the state of the art. *Studies in Second Language Acquisition*, 32(1), 1-45. <https://doi.org/10.1017/S02722263109990246>
- Browne, C., Culligan, B., & Phillips, J. (2013). *The New General Service List*. Retrieved from <http://www.newgeneralservicelist.org>
- Grendel, M., Weltens, B., & de Bot, K. (1993). Language attrition: Rise and fall of a research topic? *Toegepaste Taalwetenschap in Artikelen*, 46-47, 59-68. <https://doi.org/10.1075/ttwia.46-47.07gre>
- JASP Team. (2018). JASP (Version 0.9) [Computer software]. Retrieved from <https://jasp-stats.org/>
- Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, 24(1), 17-38. <https://doi.org/10.1080/09588221.2010.520675>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912. <https://doi.org/10.1111/lang.12079>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan.
- Weltens, B., & Grendel, M. (1993). Attrition of vocabulary knowledge. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (Vol. 6, pp. 135-156). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Poster Presentation Synopsis**Introducing a Test of the First 44 Members
of the New JACET 8000**

Tim Stoeckel stoeckel@unii.ac.jp

Dale Brown dbrown@staff.kanazawa-u.ac.jp

Most current vocabulary size or levels tests do not assess knowledge of words of the highest frequency, which are predominantly function words (Kremmel, 2016). This may be because it is assumed that these words are known by learners or that they do not hinder comprehension; because the uses of these words are heterogeneous, making it hard to determine what knowledge should be tested; or because devising a test of such words is challenging in terms of item design and writing. However, because these words provide such a large amount of coverage (Kremmel, 2016), and because many L2 vocabulary tests target beginner- and intermediate-level learners, the assumption that these words are known should be examined. Unpublished data from the New General Service List Test (Stoeckel & Bennett, 2015) indicate that some high-frequency words are less well known than others, and this particularly appears to be the case for certain function words (Table 1). Examining the assumption that the most frequent words are known is of particular importance when tests are used to estimate vocabulary knowledge in empirical research. This paper describes the development of a test, designed with Japanese learners in mind, of the first 44 entries in the New JACET 8000 list (a lemmatized frequency-based list which takes account of the uses of English within Japan; Mochizuki, 2016). These words, which are predominantly used as function words, account for 50% of the combined frequency of the items on the New JACET 8000 list. That is, in the corpora from which the list was derived, the combined frequency of these 44 items is equal to that of the other 7,956 items on the list.

Purpose

The purpose of the test is to assess whether examinees have written, receptive knowledge of one or more meaning senses or grammatical functions of the first 44 lemmas in the New JACET 8000 wordlist. For most target words, the instrument is a test of meaning-recall: the ability to recall word meaning from memory rather than identifying it from a list of choices (Schmitt, 2010). For some target words, however, the test is better conceptualized as a measure of function-recall in that test-takers must demonstrate knowledge of a word's function by offering an explanation or example (see Type 3 below).

Table 1

Item Facilities for Selected Words from the Japanese-English version of Form B of the New General Service List Test (n = 357; k = 100)

| Target Word | Words Per Million ^b | Item Facility | |
|---------------------|--------------------------------|--------------------|-----------|
| | | Value ^c | Rank |
| way ^a | 1278.59 | 1.000 | 1 |
| average | 138.23 | 1.000 | 10 |
| island | 71.96 | 0.997 | 20 |
| theory | 107.52 | 0.990 | 30 |
| decade | 78.86 | 0.974 | 40 |
| sight | 59.37 | 0.954 | 50 |
| toward | 248.23 | 0.928 | 57 |
| mayor | 17.97 | 0.919 | 60 |
| nevertheless | 51.94 | 0.866 | 66 |
| pause | 147.76 | 0.863 | 70 |
| strip | 30.28 | 0.811 | 80 |
| also | 1552.32 | 0.795 | 84 |
| brief | 68.21 | 0.713 | 90 |
| pose | 31.04 | 0.369 | 100 |

Note. Function words in bold.

^aTwelve other words, ranging in frequency from 23.79 words per million to 732.61 words per million, had item facility values of 1.000. ^bEstimated words per million reported in the New General Service List (Browne, Culligan, & Phillips, 2013; available at <http://www.newgeneralservicelist.org/>). ^cThe proportion of correct responses in the data set.

Determining which Meaning or Function to Assess

Determinations regarding which meaning senses or grammatical functions to assess were made on the basis of frequency and the notion of core meaning. First, each word's various uses were determined by consulting Google Dictionary. In some cases, Swan's (1995) *Practical English Usage* or Lindstromberg's (2010) *English Prepositions Explained* were consulted for additional information. Considering that lexical knowledge develops incrementally with multiple exposures to a word in a variety of contexts (Nation, 2013), we sometimes chose to categorize meaning senses rather coarsely. For example, Google Dictionary provides seven distinct uses of *so* as a conjunction, but these were collapsed into a single meaning sense for our purposes.

Test items were then written for the most commonly occurring meaning senses or grammatical functions for each word as determined by examining 100 randomly selected

concordance lines from the Corpus of Contemporary American English (COCA; Davies, 2008-). Since a word may have more than one significant usage, we decided to go further than many other vocabulary tests by sometimes assessing multiple uses of a word. Thus, for some words, items were written for two or even three senses or functions that occurred with similar frequency. For example, three items were written to assess knowledge of *at*. The first two probe knowledge of the word in reference to location in space and time, and the third assesses knowledge of *at* to indicate the target of perception or action (e.g., *She smiled at me*). For words that act as both function and content words (e.g., *do*, *have*), separate items were written for each unless one use was rather infrequent.

We also considered whether target words had a core meaning sense that should be assessed in addition to a more frequent but secondary use. For example, the core meaning of *have* is to possess or own something, but it occurs more frequently as an auxiliary verb. In such cases, items were written to assess understanding of both uses. By testing multiple uses of some of the words, the instrument allows learners to demonstrate partial knowledge of a word even if they do not have full knowledge of its primary uses.

Item Format

The test is divided into three sections by item type. All item types present examinees with a non-defining sentence using the target word. To increase the likelihood that even examinees of quite low proficiency would be familiar with the vocabulary in the example sentences, two steps were taken.

First, with one exception, the form of the target word used in the sentence is the lemma member that occurs with the greatest frequency in the COCA. For instance, for *say*, the example sentence uses *said*, which is much more common than other lemma members.

Second, the vocabulary used in the example sentences was carefully controlled. In many tests, this is achieved by using words of a similar or higher frequency than the target words. In this case, since each of the 44 most frequent English words was tested, this approach was obviously not possible. Nevertheless, careful lexical control was sought by almost exclusively using words from among the first 300 items in the New JACET 8000 list. The only exceptions are the use of *food*, *please*, *a.m.*, *p.m.*, and five well known proper nouns. Additionally, an effort was made to use target words only in the sentences assessing them, but due to the ubiquitous nature of the tested words, this was not always possible, and *do*, *be*, *the*, *to*, and *a* appear in items assessing other words.

Type 1

In the first item type, the target word is provided in bold followed by the example

sentence, also with the target word in bold. Examinees are asked to provide an L1 translation, paraphrase, or explanation of the target word. Here is an example prompt for *she*:

*she: **She** knows a lot.*

Type 2

For some target words, it is difficult or impossible to provide a one-to-one translation or paraphrase in Japanese. An example is the modal verb *do*, which in negative constructions manifests as a verbal inflection in Japanese rather than as a separate word form (e.g., *run* and *do not run* are *hashiru* and *hashiranaï*). For such words, examinees are presented with an example sentence and instructed to translate the entire sentence. Following Stoeckel, Ishii, and Bennett (2018), answers are evaluated as to whether the meaning communicated by the target word is understood even if non-target words are mistranslated. This, together with the use of very high-frequency words and simple grammatical structures in the example sentences, reduces the risk of conflating target word knowledge with other types of knowledge. Here is an example for *do*:

The students don't like me.

Type 3

For a few target words, it is difficult to assess comprehension with either of the item formats above. An example is the use of *will* to talk about the future. In Japanese a single verb tense can express both habitual and future actions, so a translation of a sentence using future *will* (as in item type 2) would be indistinguishable from a sentence using the simple present tense. In such cases, an example sentence follows a question in Japanese to elicit whether the meaning of the target word is understood. For example, for *will* the test item is:

下の文では「will」は何を伝えますか？

[*In the sentence below, what information is communicated by the word "will"?*]

Takeshi will come to my house.

The Test

In total, the test consists of 65 items across three sections as shown in Table 2. Type 2 items were placed first because this item type does not draw examinees' attention

to the target word, which is of importance given that for some words another use is tested later. Likewise, item type 3 was last because some of these questions directly ask learners to think about the word being tested.

Table 2

Breakdown of Test by Sections

| Section | Item Type | Number of Items |
|---------|-----------|-----------------|
| 1 | 2 | 34 |
| 2 | 1 | 26 |
| 3 | 3 | 5 |

Conclusion

The primary purpose of the test described above is to ascertain whether knowledge of very high-frequency words, many of which are function words, can be assumed for Japanese learners of English, especially those of low proficiency. If we find this to be the case, it would provide validation evidence for (a) levels and size tests whose design implicitly assumes knowledge of such words and (b) research based on the use of such tests. If not, test items assessing knowledge of very high-frequency words ought to be added to existing and future tests that are used with low-proficiency learners.

References

- Browne, C., Culligan, B., & Phillips, J. (2013). *The New General Service List*. Retrieved from <http://www.newgeneralservicelist.org>
- Davies, M. (2008-) *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present* [Corpus]. Available online at <https://corpus.byu.edu/coca/>
- Kremmel, B. (2016), Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, 50, 976-987. <https://doi.org/10.1002/tesq.329>
- Lindstromberg, S. (2010). *English prepositions explained*. Amsterdam: John Benjamins Publishing Company.
- Mochizuki, M. (2016). *JACET 8000: The new JACET list of 8000 basic words*. Tokyo: Kiriara.
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.
- Stoeckel, T., & Bennett, P. (2015). A test of the New General Service List. *Vocabulary*

Learning and Instruction, 4(1), 1-8. Retrieved from <http://vli-journal.org/wp/vli-v04-1-2187-2759/>

Stoeckel, T., Ishii, T., & Bennett, P. (2018). Is the lemma more appropriate than the flemma as a word counting unit? *Applied Linguistics*. Advance online publication. <https://doi.org/10.1093/applin/amy059>

Swan, M. (1995). *Practical English usage* (2nd edition). Oxford: Oxford University Press.

Poster Presentation Synopsis**Motivating Low-Achievement Learners to Prepare for
Weekly Vocabulary Tests:
The Come-Back Vocabulary Test**

Magda Kitano magda@shonan.bunkyo.ac.jp

Katsuhiko Chiba ch1ba@shonan.bunkyo.ac.jp

Theoretical Framework

While 6,000 to 9,000 words is a recommended goal for second language learners of English, high-frequency vocabulary is particularly important. As these words make up a large percentage of both spoken and written text, a lack in fluency of high-frequency words is a serious obstacle to comprehension and language production (Nation, 2013). However, while first-year university students in Japan know a range of only 1,700 to 2,300 words after six years of English instruction (Barrow, Nakanishi, & Ishino, 1999), studies such as Browne and Culligan (2008) have found that there are large gaps in their knowledge of the most frequent words. Combining this problem with the limited class time provided in Japanese universities, Browne (Wadden, Browne, & Nation, 2019) recommended the usage of out-of-class time for vocabulary study and that instructors set regular vocabulary-learning goals such as a weekly list of words to learn. Robb (2019) also indicated that out-of-school study in all aspects of English learning is necessary under the Japanese university system in order for students to reach their TOEIC goals.

Compounding the problem is a general lack of motivation among Japanese students to put in time for study. Robb (2019, p. 116) noted that while there are a few highly motivated students in Japanese universities, for the “vast majority” a teacher should expect them to try to get away with as little work as possible. Snyder (2019, p. 137) wrote that although students lacking motivation to learn English is a stereotype, such students are in fact a “frequent source of frustration” for university instructors in Japan, with students “at best . . . dutiful but barely participatory; at worst, they are indifferent to homework and exams, often tardy, or absent altogether.” Teachers of remedial learners face even more challenges. Kiyota (2009) noted a motivational gap between remedial and regular learners in Japanese university English classes. He concluded that different teaching approaches are necessary for these two types of learners. This situation is a particular challenge for vocabulary learning, which requires multiple retrieval events and spaced repetitions for successful acquisition (Nation, 2013).

The current study was an attempt to motivate near-remedial students to spend time

outside of class preparing for weekly vocabulary tests by personalizing the testing system. If successful, it could be a stepping stone to better study habits across the curriculum for students, and high-frequency words learned in the process would benefit students by making English more accessible.

Sample

The subjects were first-year students at a mid-level university in Kanagawa, Japan, all majoring in international studies. At the beginning of the semester all incoming first-year students took a test of the 1,000 most frequent words of the New General Service List (NGSL; Browne, Culligan, & Phillips, 2013), were instructed on the importance of high-frequency words to their English abilities, and were assigned the task of identifying unknown words in the first 1,000 of the NGSL and learning those for a test at the end of the semester.

Required English classes were divided into six levels for first-year students as determined by scores on the Computerized Assessment System for English Communication (CASEC), with two classes of students at each level. The subjects of this study were those of the two lowest-level classes, with CASEC scores ranging from 219 to 405 ($M = 341.68$, $SD = 43.88$) for class A, and 122 to 315 ($M = 260.06$, $SD = 59.82$) for class B. These scores are equivalent to CEFR levels B1 and B2 (“Data & Information,” n.d.). Of the total of 44 students in these classes, weekly test score data were analyzed for only those whose CASEC scores were available ($N = 36$).

Procedure

In order to aid students of these lowest-level classes in reaching the program-wide goal of mastering the first 1,000 words of the NGSL during the spring (first) semester of the 2018 school year, students were assigned weekly English-Japanese bilingual lists of ten words to learn. These words were chosen from those unknown by many in the initial test, and also included commonly unknown words from previous years. Weekly tests required students to write the target words from cues of the Japanese equivalents.

When the students received the results of the first weekly test, they were told that all words that were marked wrong on this test would appear on the following week’s test along with the new words for that week. Subsequent weekly tests were personalized, with a test of the new words of the week common to all students on one side of the paper, and each student’s “come-back” words from all previous tests on the other. Once a word was correctly supplied by the student, it no longer reappeared on future come-back tests. Grades for student assessment were determined as the percentage of correct responses on the entirety of the test (both sides of the paper).

This system was made feasible by utilizing the software FileMaker Pro (version 16) to create a database that kept track of test items, students, and test results. After indicating *accurate* or *mistaken* in the database for each word on the tests, subsequent personalized tests could be printed automatically. Personalized test marking also became feasible because the student's current test answer key could be viewed on-screen.

Six weekly lists were introduced and tested. For the two weeks following, only come-back words were tested. Students with no come-back words returning to them were rewarded with full scores for those weeks. Finally, in an end-of-semester questionnaire, students were asked their impressions of this testing method and whether they thought this program should be continued in future classes. Of the 44 students, 41 were present on the questionnaire day.

Preliminary Results

Looking at the weekly test scores of new target words only (i.e., not including students' come-back words), scores rose over the six weeks (Figures 1 and 2). In paired-samples *t*-tests comparing performance on the first and last weekly tests, significant improvement was seen. In class A, the first test had a mean of 7.05 ($SD = 4.31$), and the final test a mean of 8.92 ($SD = 1.71$), $t(18) = 2.50$, $p = .022$, $d = .57$. In class B, the first test had a mean of 6.05 ($SD = 3.60$), and the final test a mean of 8.88 ($SD = 2.21$), $t(16) = 3.11$, $p = .007$, $d = .95$.

Illustrating how these results differ from the usual situation in classes at this level, Figure 3 shows test scores from seven weeks of a 2017 class, also the lowest of six levels. Students in this class were similarly tested on ten words weekly from the first 1,000 words of the NGSL, but without the come-back method being used. While high scores are seen throughout the period, many students peak and then return to zero, indicating sporadic test preparation. In comparison, Figures 1 and 2 with the come-back method show very few zero or low marks once the program started up, while most marks cluster at the top. As to improvement between the first and last tests, a paired-samples *t*-test of the 2017 data did not indicate significant improvement (week one $M = 5.47$ ($SD = 4.40$), week seven $M = 6.26$ ($SD = 3.80$), $t(18) = 0.67$, $p = .508$, $d = .19$.)

Questionnaire answers were in general positive about the method, with 37 students answering that they thought the come-back system should be used in future classes, and two answering that it should not (2 abstentions). In a free-answer question, 29 comments were positive, some glad that the system pushed them to study when they normally would not have, and some appreciating the chance to try again to get words right that they had missed. Five comments were negative, some disliking how long their come-back word lists grew.

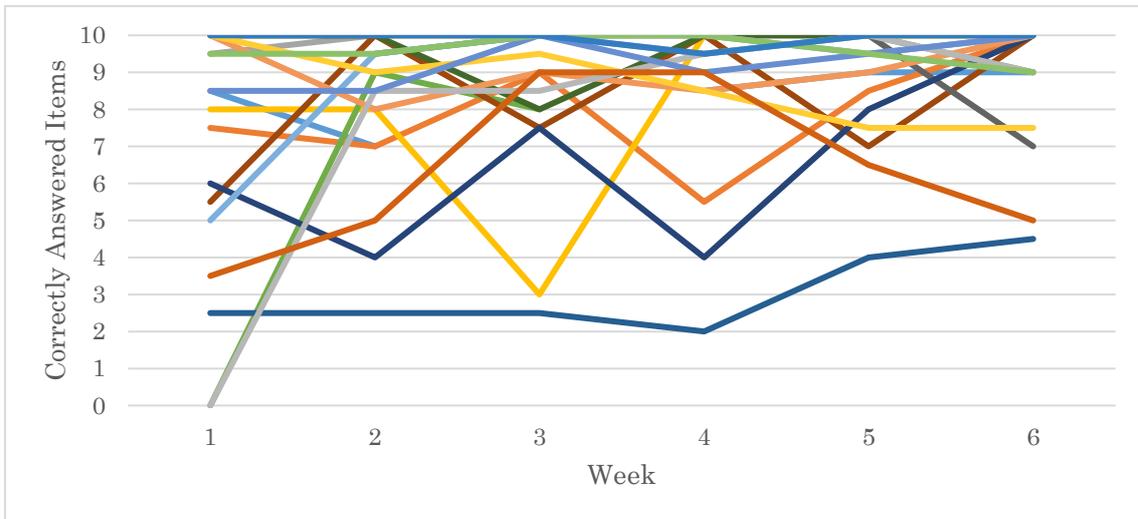


Figure 1. Class A student scores for weekly vocabulary tests. Scores include only the list of ten new words assigned for the week. One half point was given for misspelled words.

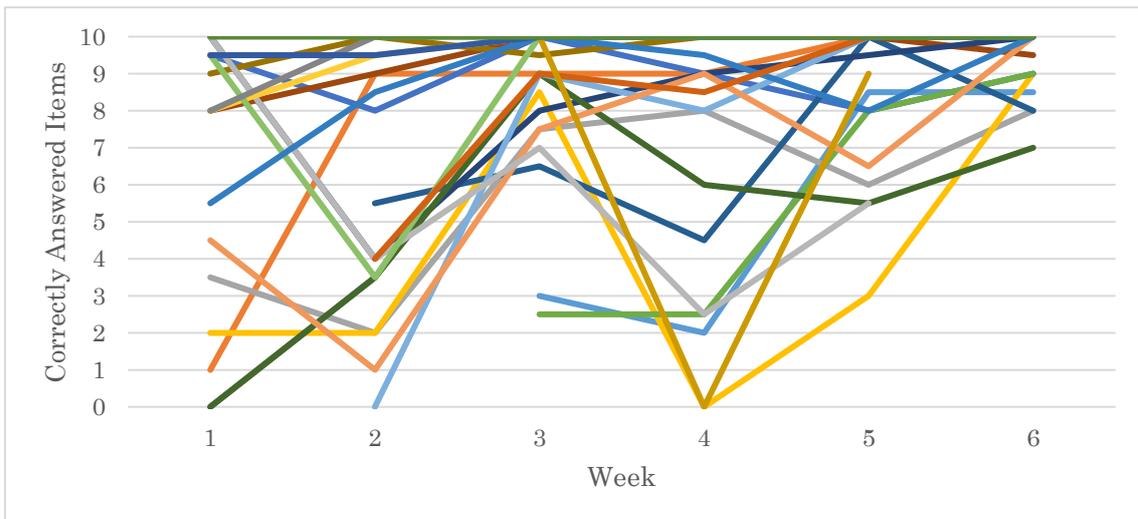


Figure 2. Class B student scores for weekly vocabulary tests. Scores include only the list of ten new words assigned for the week. One half point was given for misspelled words.

Preliminary Conclusions

In a group of students who were not in the habit of completing homework assignments or preparing for vocabulary tests, the come-back method was seen to encourage most of the students to study before tests, with a goal of getting as many correct as possible. The prospect of an increase in the amount of words one must study seemed to be a key aspect of this system, and success was rewarded immediately by correctly-answered words being taken off students’ come-back lists.

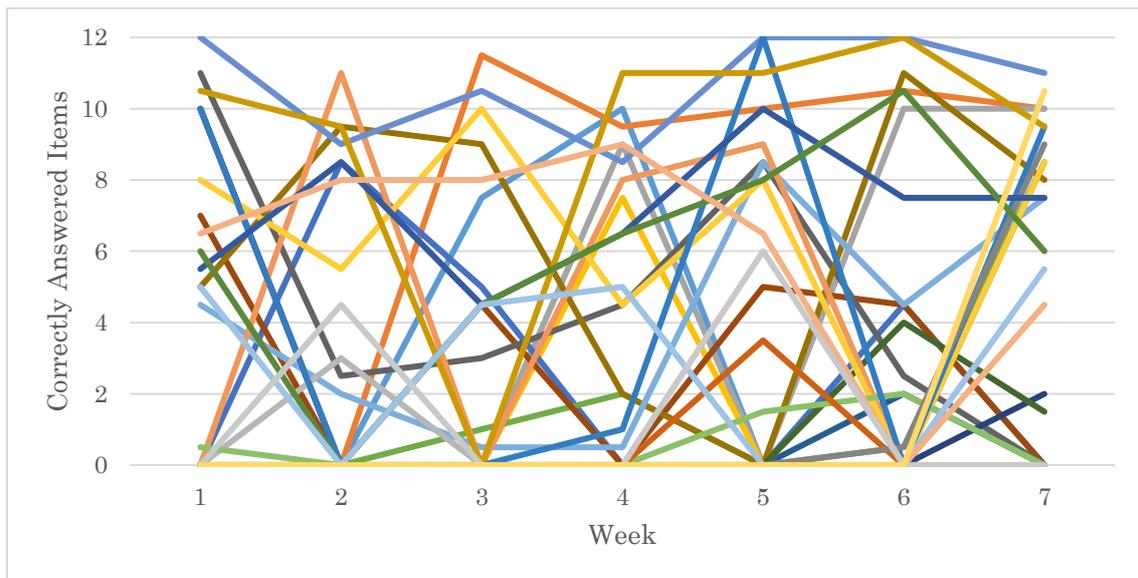


Figure 3. Student scores for weekly vocabulary tests, 2017 class of second year students. In addition to ten weekly vocabulary words, tests included two bonus point questions concerning textbook content, resulting in a total of 12 possible points.

Future Directions

This study only looked at the impact of the come-back method on weekly vocabulary scores; it did not investigate whether words were being acquired. Future research on this teaching practice ought to involve a control group given weekly lists and tests without the come-back aspect, and would need to test for target word knowledge retained over time.

References

- Barrow, J., Nakanishi, Y., & Ishino, H. (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System*, 27, 223-247.
- Browne, C., & Culligan, B. (2008). Combining technology and IRT testing to build student knowledge of high frequency vocabulary. *The JALT CALL Journal*, 4(2), 3-16.
- Browne, C., Culligan, B., & Phillips, J. (2013). *The New General Service List*. Retrieved from <http://www.newgeneralservicelist.org>
- Data & information. (n.d.). Computerized Assessment System for English Communication. Retrieved from <https://global.casec.com/data/>
- Kiyota, Y. (2009). Motivation of remedial EFL learners (A case study of Japanese college EFL learners). *リメディアル教育研究* [The Japan Association for Developmental Education] 4(2): 173-179.

- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Robb, T. N. (2019). Homework in the Japanese university classroom. In P. Wadden & C. C. Hale (Eds.), *Teaching English at Japanese universities* (pp. 115-121). New York, NY: Routledge.
- Snyder, B. (2019). Creating engagement and motivation in the Japanese university language classroom. In P. Wadden & C. C. Hale (Eds.), *Teaching English at Japanese universities* (pp. 137-143). New York, NY: Routledge.
- Wadden, P., Browne, C., & Nation, P. (2019). Teaching and learning vocabulary in the Japanese university. In P. Wadden & C. C. Hale (Eds.), *Teaching English at Japanese universities* (pp. 84-96). New York, NY: Routledge.



SIG News

Upcoming Events



October 12, 2019
Waseda University
Nishi-Waseda Campus,
Building 63

Hosted by:



Center for English Language
Education in Science and Engineering
Waseda University

Proudly sponsored by:



Session 1: Learning

Discussant

Irina Elgort
Victoria University of Wellington

Speakers

Chie Ogawa
Kyoto Sangyo University
Darrell Wilkinson
Tokyo Woman's University
Haidee Thomson
Hokusei Gakuen University
Victoria University of Wellington
Michael Holsworth
Kyoto Sangyo University

Session 2: Vocabulary & CALL

Discussant

Tatsuya Nakata
Hosei University

Speakers

Clint Denison & Imogen Custance
Mukogawa Women's University
Kwansei Gakuin University
Louis Lafleur
Ritsumeikan University
James Rogers
Meijo University
Andrew Obermeier
Kyoto University of Education

jaltvocab.weebly.com/symposium

2019 JALT Vocabulary SIG Symposium Call for Poster Presentations

Waseda University, Nishi-Waseda Campus, Building 63, Tokyo
October 12, 2019

In order to ensure a strong audience, no symposium sessions will run concurrent to poster presentations. Proposals for poster presentations will be accepted under two general categories:

1. Ongoing vocabulary-related research
2. Vocabulary teaching and learning in practice

Submission:

Please e-mail 1) your name, 2) presentation title, and 3) a brief outline of your proposed topic (approximately 300 words), to <jaltvocab[at]gmail.com>

Deadline: Sunday, August 31, 2019

Abstracts should follow one of the two following formats:

Ongoing research format:

- Background
- Aims
- Methods
- Preliminary results
- Preliminary conclusions
- Future directions

Teaching and learning format

- Theoretical framework
- Procedure
- Preliminary results
- Preliminary conclusions
- Future directions

Accepted poster presenters will be invited to submit written, extended synopses of their presentation for publication (500 – 1500 words) in an upcoming issue of VERB.

VERB Call for Papers

The VERB welcomes submissions related to vocabulary research and education.

Short papers are peer reviewed and may require rewriting and resubmission for acceptance. They must not exceed 1500 words, excluding references, tables, and titles. Short papers fall into the categories of completed research, ongoing research, and teaching and learning in practice.

Other submissions encouraged are classroom activities related to vocabulary, book reviews, opinion pieces, and event reports and commentary. All submissions are expected to adhere to APA 6th edition formatting guidelines.

Deadline for next issue: **August 31, 2019.**

For submissions and all correspondence: <jaltvocabsig.verb@gmail.com>

Latest information: <https://jaltvocab.weebly.com/publications.html>

The following are guidelines for short paper submissions (please include these sections):

| Completed research: | Ongoing research: | Teaching and learning in practice: |
|----------------------------|--------------------------------|---|
| * Background | * Background | * Theoretical framework |
| * Aims | * Aims | * Teaching context |
| * Methods | * Methods | * Procedure |
| * Results | * Sample | * (Preliminary) Results |
| * Conclusions | * (Preliminary) Results | * (Preliminary) Conclusions |
| * Future directions | * (Preliminary) Conclusions | * Future directions |
| | * Future directions | |

**If you are thinking about submitting, but your article doesn't fit into one of the above categories, please email us at the above address and let us know what you would like to submit and we can work it out.

***Vocabulary Learning & Instruction* Call for Papers**

The Vocabulary SIG's *Vocabulary Learning and Instruction* (VLI) journal is calling for submissions for an upcoming issue. Submissions will be published online upon acceptance, and combined into an issue later this year.

VLI accepts long-form research papers (2000-7000 words) and brief reports, summaries, and commentaries (2000-3000 words) related to vocabulary acquisition, pedagogy, assessment, and lexical networks.

As an open journal, content is indexed on Google Scholar and made freely available on the internet without paywalls. Authors are free to also make their work available on sites such as academia.edu and researchgate.

All submissions are subject to a 2-step peer-review process:

A) Editors review manuscripts to ensure basic requirements are met, and that the work is of sufficient quality to merit external review. This process typically takes 1-2 weeks, at which point authors are informed of the outcome.

B) Submissions which meet these requirements are sent out for blind peer review by 2-3 experts in the field. This process takes approximately 1-2 months. Following external review, authors are sent copies of external reviewers' comments and notified of decisions (accept, accept pending changes, revise and resubmit, or reject).

Please see <http://vli-journal.org/submissions.html> for details

Vocabulary SIG 2019 Research and Conference Grants FAQ

1. What grants are available?

There are two categories of grants, *research* grants and *conference* grants. There are separate application forms for each. Up to three grants of either 50,000 or 100,000 yen may be awarded in 2019 depending on the number of viable applications.

2. Who can apply for the grants?

Grants are available for JALT Vocabulary SIG members who do not already have access to institutional assistance and research grant funding from their place(s) of employment. The grants are specifically designed for instructors who do not have access to institutional assistance and research grant funding from their place of employment. This includes part-time instructors or those working from the kindergarten to tertiary levels, as well as individual instructors working for themselves, companies, language schools. Part-time instructors who are enrolled as part-time or full-time graduate students are also eligible. Grant awardees must have a Japanese bank account.

3. What is the deadline for applying for grants in 2019?

Applicants must submit their applications by 23:59 (JST) on Tuesday, September 10th, 2019. Applicants will then be informed of the results after the vetting period is completed on or around September 20th, 2019.

4. Where can I get more information?

You can find the complete grant information on the Vocab SIG website at <https://jaltvocab.weebly.com/grants.html>

If you have any other questions, please contact the Vocabulary SIG Research Grant Chair, Brandon Kramer, at JALTvocabSIGgrants@gmail.com