

Vocabulary: Describing the College Entrance Gap in Japan

Paul Sevigny¹ and Lance Stilp²

Abstract

This study sought to determine the vocabulary gap for domestic Japanese learners of English prior to college entrance. Teacher-researchers analyzed the 1K and 2K lists from the General Service List (GSL) (West, 1953) and created two separate sets of Yes/No tests for these respective lists. The 2K list was divided randomly into nine Yes/No tests of approximately 110 items each, and each test was given to different groups of approximately 37 pre-intermediate English learners (total N=334) at a medium-sized Japanese university in the fall of 2013. The 1K list was similarly divided into eight Yes/No tests and given to the same-sized groups (total N=298) in the fall of 2014. Test-takers claimed to know about 47% of the 2K list and 72% of the 1K list items after correcting for false alarm rates of 8 and 9% respectively. The results support Browne's (1998) claim that high school English curricula do not develop adequate depth of knowledge of the most frequent English vocabulary among high school English learners in Japan. This has led to creating a list of the "Forgotten 400" unfamiliar 1K and 2K words, which is provided along with the item facilities (IF) for each of these words.

Key Terms: vocabulary, Yes/No tests, diagnostic tests, needs assessment, curriculum design, GSL, NGSL

1. Introduction

Browne (1998) has stated that the extent to which Japanese high school education focuses on college entrance tests has fostered the use of high school textbooks that present overly-complex lexis—too much vocabulary from low-frequency bands—according to corpus-based vocabulary profiles. He has concluded, "Instead of forcing students to spend inordinate amounts of time memorizing so many low frequency words to deal with their assigned texts, it would seem a wise and prudent step to first make sure that students get control of the high frequency words and University Word List vocabulary" (Browne, 1998, p. 10). With this idea in mind, teacher-researchers aimed to identify high-frequency English vocabulary items that are generally unknown to our population in the pre-intermediate level of the English language program, the standard college entrance level at our university. An additional motivation came from implementing Cih'i's (2013) Word Engine spaced-repetition software (SRS) system, which uses a statistical system to check the learner's vocabulary size and then prescribes an individual course of the most important words for each learner. Previously, the approach to integrating vocabulary into course design was to adopt and adapt the common course lists from a textbook and to test learners on the bold-faced words chosen by a publisher. Moving from a one-size-fits-all vocabulary system to individually-differentiated flashcard study would improve our curriculum, but the Word Engine system did not provide full information about each learner's prescribed vocabulary. We were concerned that these two factors (lack of information and student tendency to overreach) might perpetuate a situation in which learners lack knowledge of highly frequent vocabulary and thus continue to rely on translation due to a persistent lack of coverage. Therefore, we were particularly interested in closing major gaps in high-frequency vocabulary that first-year students might have. The results of this study could also help to shape the vocabulary program design used in pre-enrollment schooling, a special program to help learners accepted to our university study effectively between high school graduation and university matriculation.

¹Tenured lecturer, Ritsumeikan Asia Pacific University (APU), Beppu City, Oita, Japan

e-mail:sevignyp@apu.ac.jp

²Lecturer, Ritsumeikan Asia Pacific University (APU), Beppu City, Oita, Japan

e-mail: ljstilp@apu.ac.jp

2. Literature Review

There are four essential areas of literature in vocabulary learning that form the foundation of this study. The first area is the use of Yes/No tests for diagnostic vocabulary testing. The second is the comparison of the Yes/No tests in this study and a closely related study by Sevigny and Ramonda (2013). The third is the attempt to estimate Japanese college students' average vocabulary size, which demonstrates the need for the present study and more detailed description of gaps in learner knowledge. The last essential area of literature is the interconnection of word frequency lists and the concept of coverage.

The first area of literature essential to this study is the use of Yes/No tests (sometimes called checklist tests) for teaching and researching vocabulary. A Yes/No test simply presents a word without context and asks the learner to indicate whether the word is known or not. Paper versions may require the learner to circle or check the known words. Computer-based versions typically have the learner click on a radio button. In order to ensure test-takers are making sound judgments, the use of nonwords like *wuggy* were introduced to identify whether given test-takers were incorrectly reporting knowledge of a word. These nonwords yield a false alarm rate: the average number of times learners claim to know nonwords. There have been numerous studies that support the general use of Yes/No tests for curricular development and placement tests (Shillaw, 2009; Mochida & Harrington, 2006). Nation (2008) and Read (2007) have endorsed the use of Yes/No tests to assist with student placement and to supplement the use of the Vocabulary Levels Test (VLT). While the VLT provides general data with reference to vocabulary frequency bands, Read (2007) advised that Yes/No tests can assess vocabulary size with respect to specific lists, thus helping programs to produce context-specific vocabulary tests. Furthermore, because of its simplicity, the Yes/No test is "informative and cost effective" (Read, 2007, p. 113). Not only do the studies above support the validity of the Yes/No format, but Shillaw (1996, p. 7) used Rasch analysis to show the use of nonwords had little effect on the performance of his learners.

Sevigny and Ramonda (2013) and this study employed the same Yes/No test format and methodology to impact the vocabulary layer of a curriculum design but differ in important ways. These differences can best be described by comparing three key elements: the approach to vocabulary study in the teaching context, the scope of the curriculum design, and the local impact of each study. First, with regard to the approach to vocabulary study, Sevigny and Ramonda (2013) used Yes/No tests in a course where a common course list of 240 items had been adopted from a textbook and had been turned into summative vocabulary test items. The course took a one-size-fits-all approach to guiding learner vocabulary study and assessment. On the other hand, in the context of this study, individually differentiated flashcard study was the dominant mode, aimed towards adding 1000 words to each learner's vocabulary per semester, especially for TOEFL study. Second, with regard to the scope of the curriculum design, Sevigny and Ramonda (2013) used Yes/No tests within one pre-intermediate course to determine the vocabulary that should actually appear in summative vocabulary assessments. The current study has a larger scope, aiming to help define boundaries between pre-enrollment and first-semester individual flashcard study. Finally, the local impacts of these studies differ. Sevigny and Ramonda (2013) resulted in better formative vocabulary testing with reference to specific textbook word lists and also helped move a program away from a one-size-fits-

all approach of only testing bold-faced vocabulary from a common course list. This made students responsible for all key words in their intensive reading book in addition to newly implemented, individualized vocabulary study. The local impact of the current study will be the development of supplementary, computer-based vocabulary packs that can be assigned to learners with vocabulary deficits in high-frequency bands.

The third area of literature relates to the varying attempts to estimate the size of Japanese English language students' vocabulary. This is important because measures of vocabulary size can be inflated, or even if accurate, obscure gaps in learner knowledge of high frequency vocabulary. Barrow, Nakanishi, and Ishino (1999) estimated that first-year Japanese college students possess a vocabulary of around 2,300 words. These teacher-researchers used checklist (Yes/No) tests with nonwords to estimate vocabulary size. This line of research has been updated recently by Mclean, Hogg, and Rush (2014), using Nation and Beglar's (2007) Vocabulary Size Test (VST). The VST is an 80-item multiple choice test that represents the first 8,000 word families of English. These teacher-researchers found an average score of 39.39 out of 80, which theoretically represents knowledge of 3,939 word families. However, the authors concede that the multiple-choice nature of the test has inflated the results to some degree. Gyllstad, Valkaite, and Schmitt (2015) report that process-of-elimination guessing can inflate multiple-choice vocabulary test scores from between 11% to 26% (p. 289). As a worst-case scenario, it appears that there could be a 1000 word discrepancy in the results reported by Mclean et. al. (2014). Another measure of vocabulary size used in Japan over the past ten years is Word Engine's V-check. According to data published by Lexxia, the maker of Word Engine (Cih, 2013, p. 2), a learner with a paper-based TOEFL score of about 420, the average for many students completing pre-intermediate English at our university, would know about 3,500 words. This is similar to the preliminary findings by Mclean et al. (2014), which did not provide information about participants' proficiency levels. Relying solely on information based on average vocabulary size becomes very problematic for administrators developing course curricula. Whether first year students have a vocabulary size of 2,000 or 3,000 words, perhaps a more important question is whether these learners know the most frequent 2,000 to 3,000 words. This study will begin to address both this gap in the learners' vocabulary knowledge and in published research.

The last essential area of literature, the relevance of frequency bands, cannot be conveyed without a review of *word frequency lists* and the concept of *coverage*. The General Service List (GSL) published by West (1953) consists of 3,372 headwords ranked by frequency of occurrence. Nation (2008) estimates that 80% of most texts are comprised of the most frequent 2,000 word families of the GSL. Browne, Culligan, and Phillips (2013) have introduced a revision of the GSL called the New General Service List (NGSL), a list of 2,801 words that comprises 92% of most texts. *Coverage* refers to the percentage of the total words in a given passage that a given learner knows receptively. The most commonly cited coverage needed for adequate comprehension for L2 readers is between 95% (Laufer, 1989) and 98% (Hu & Nation, 2000). Determining a subset of the GSL or NGSL that is unfamiliar to a group of learners would provide a useful and relevant list for deliberate study because this knowledge would boost the coverage that a given learner brings to a given reading.

3. Context

The current research was situated in the pre-intermediate English level of a large EFL program at a Japanese university: the same context as Sevigny and Ramonda (2013). The courses in this level are divided by skills into a two-credit reading and vocabulary course and a four-credit listening, speaking, writing, and grammar course. The data collection for both studies was carried out in the two-credit reading and vocabulary course. In the fall of 2013, learners were placed into the pre-intermediate level with the paper-based TOEFL test. By the fall of 2014, the Cambridge English Placement Test had replaced the TOEFL.

Starting in the spring of 2014, once learners were placed into their level, the Word Engine V-check was used to place learners within the Word Engine spaced repetition software (SRS) system. Word Engine added easy access to flashcard study via computers or smart devices for the reading and vocabulary course. This individualized vocabulary program accounted for 30% of each pre-intermediate student's grade. Students were required to complete 600 correct responses (CR) per week for 15 weeks. If a student did not complete the 600 CR then he or she lost 2% of the total grade for that week. In order to keep the data consistent with that collected in 2013, the 1K band Yes/No tests were given in the first week of instruction in the fall of 2014, before students started using this system.

3.1 Participants

There were two sets of participants in our study. The first study tested 334 university students of pre-intermediate English in the fall of 2013, and the second tested 298 in the fall of 2014. The mean paper-based TOEFL score for both of these cohorts was approximately 415 at the end of the semester. While the majority of these learners were from Japan, the population is also comprised of some learners from Korea and China. Approximately 37 participants in the fall 2013 study were not Japanese and listed as international students. Groups were formed by assigning two classes with approximately 20 students each to take one Yes/No test. In 2013, students were randomly assigned to gender-balanced groups. In the fall of 2014, one class was specially designated as an above-average class, but students in all other classes were assigned randomly to gender-balanced groups.

4. Research Issues

In the spring of 2013, informal action research using Nation's (2008) Vocabulary Levels Test (VLT) showed that about 200 or 300 words on the 2K GSL were potentially unknown to our pre-intermediate population. We recollected Read's (2007) recommendation for using Yes/No tests to supplement the VLT. It was a natural progression to apply the Yes/No Test to the 1K and 2K bands of the GSL, but to test so many items would require breaking these sets of 1,000 items into random sets of about 100 items to avoid test fatigue. This would be at the upper limit of Nation's (2008) recommendation. Since the 2K list appeared to provide more opportunity for charting unfamiliar words, this band was scheduled for diagnostic testing in the fall of 2013. The 1K list was scheduled for testing in the fall of 2014.

The following are the proposed research questions:

1. To what extent are pre-intermediate learners familiar with the 2K band of the GSL?
2. To what extent are pre-intermediate learners familiar with the 1K band of the GSL?
3. If these results are profiled using the NGSL, which words could be assigned to pre-enrollment learners for review prior to college entrance?

5. Methods

5.1 Procedure

Two separate sets of Yes/No tests were constructed in order to test the most commonly known words at our institution. The first set was an adapted 2K GSL list containing 990 words divided into nine tests. These tests were given to nine groups of pre-intermediate level English students on December 10, 2013. The second set was a similarly adapted 1K GSL list containing 880 words, divided into eight different tests. These tests were given to different incoming pre-intermediate level English students at the very beginning of the semester on October 16, 2014 before any instruction had been given. All tests were assigned to approximately 40 students. The tests were administered electronically through the Blackboard 9.1 learning management system. Teachers were instructed to provide only ten minutes to complete the test and were given a PowerPoint file containing detailed test instructions in both English and Japanese (Appendix A).

Within the 1K and 2K lists, words were ordered randomly, so they would not appear in alphabetical order. Slight variations were made in the order if two adjacent words were too similar in spelling and/or meaning. Each test had approximately 110 correct question items, in which students either clicked “yes” if they knew the word or “no” if they did not. The same 15 nonwords were evenly spaced throughout each test to keep learners engaged and to assess the false alarm rate, which is the rate learners might incorrectly confirm knowledge of a word. Including nonwords, there were about 125 items per test.

First, we adapted the 2K and 1K lists by removing 101 words from the 2K GSL that we deemed to be already known (e.g. *rice*, *telephone*). The words that have been incorporated into Japanese as loanwords were also eliminated (e.g. *hotel* becomes *hoteru*). These types of words called *gairaigo* might have contributed to an interference from the L1, compromising the validity of the data. For consistency, we changed all spelling conventions into American English. Finally, a total of 91 randomly selected words were added from the 2K NGSL to replace these excluded words. Both tests also used plausible nonwords as a means to check the validity of test-takers’ judgments (Meara & Buxton, 1987). From a selection of 99 plausible nonwords, a total of 15 nonwords were selected (Appendix B). When creating the 1K GSL list in the following year (2014), we similarly removed words we thought our students already knew and created similar test sizes of 110 test items plus 15 nonwords.

5.2 Scoring

This study used the same three scoring procedures as in Sevigny and Ramonda (2013):

1. Average test score: the average percent of the real words the learners claimed to know
2. Item facility: the percentage of learners who claimed to know that item
3. False alarm rate: the total number of false alarms made by all participants divided by the total number of nonwords presented on the three forms (p. 705)

The Blackboard 9.1 system uses radio buttons that students click to register their responses. Responses are automatically scored, providing the number of words known for each student. Blackboard 9.1 yields score reports for individual learners, group averages, and individual test items, including the percentage of learners answering each item correctly.

6. Results

All participants' responses are automatically saved in Blackboard 9.1, and there were no participants whose test results had to be removed from the study for test malfunctions. There were occasions where no answer was given by some students, but these were very few and not considered to have significantly affected results.

The results for the 2K tests and the 1K tests are reported in Tables 1 and 2 respectively. The test scores are reported in both raw numbers and as percentages along with the corresponding frequency with which the groups (pairs of classes) correctly rejected nonwords. From Table 1, pre-intermediate level students on average claimed to recognize 55% of the adapted 2K GSL list and correctly rejected about 13.8 of 15 nonwords across all groups. For the 2K tests, the average false alarm rate was about 8% (Appendix B). Read (2007), after examining multiple studies of complex correction formulas, stated the following:

It appears that in practical testing situations, if it can be assumed that the test-takers are honestly reporting most of the time whether they know the target words or not, a simple calculation such as the number of Yes responses to real words minus the number of Yes responses to nonwords yields a reasonably valid measure of vocabulary size.

(Read, 2007, pp. 110-111)

Thus, according to Read's instructions above, subtracting 8% from the average test score would adjust the 55% result to about 47% of the adapted 2K list being known to these students.

Table 1

2K Test Results by Group (N=334) Test Date: December 10, 2013

Test Group	Average number of known words (k = 110)	Average number of nonwords correctly rejected (k = 15)
Group 1 (n=45)	64.44 (59%)	14.4 (96%)
Group 2 (n=36)	54.15 (49%)	14.1 (94%)
Group 3 (n=40)	64.77 (59%)	13.6 (90%)

Group 4 (n=37)	59.67 (54%)	14.0 (93%)
Group 5 (n=37)	58.23 (53%)	14.2 (95%)
Group 6 (n=34)	67.60 (62%)	13.0 (87%)
Group 7 (n=33)	64.14 (58%)	13.7 (91%)
Group 8 (n=38)	56.69 (52%)	13.7 (91%)
Group 9* (n=34)	53.18 (49%)	13.8 (92%)
Total Average	60.32 (55%)	13.8 (92%)

*Group 9 (k=109)

The false alarm rates for all nonwords for both the 2K and 1K GSL tests are reported in Appendix B. Both sets of tests used identical nonwords (Cobb, 2010) with the exception of *runster* (used in the first 2K test) and *kiley* (used in the second 1K test). The average false alarm rate for nonwords on the 2K set of tests was 7.75%. The average false alarm rate for nonwords on the 1K set of tests was 9.34%.

Table 2 indicates that incoming pre-intermediate level students claimed to recognize 81% of the adapted 1K GSL list. They also correctly rejected 13.6 of 15 nonwords. According to Read's (2007) adjustment, subtracting 9% for the false alarm rate would result in an average test score of about 72% of 1K GSL words being known to pre-intermediate learners in the first week of instruction.

Table 2

1K Test Results by Group (N=298) Test Date: October 16, 2014

Test Group	Average number of known words (k = 110)	Average number of nonwords correctly rejected (k = 15)
Group 1 (n=40)	88.68 (81%)	13.9 (92%)
Group 2 (n=37)	89.44 (81%)	13.4 (89%)
Group 3 (n=40)	91.91 (84%)	13.4 (89%)
Group 4 (n=39)	90.78 (83%)	13.9 (92%)
Group 5 (n=30)	85.95 (78%)	14.2 (95%)
Group 6 (n=40)	86.20 (78%)	13.0 (86%)
Group 7 (n=35)	88.77 (81%)	13.3 (89%)
Group 8* (n=37)	89.69 (84%)	13.8 (92%)
Total Average	88.93 (81%)	13.6 (91%)

*k=107 for Group 8

These results support the notion that there are large gaps in the receptive vocabulary knowledge of pre-intermediate Japanese university students with respect to highly frequent vocabulary, especially regarding the 2K GSL. Each item on the

adapted 2K and 1K GSL list now has a corresponding item facility (IF) determined as a result of these test administrations. For example, the word *severe* is in the 2K frequency band, and 15.4% of learners claimed to know it. Thus, the item facility is 15.4. On the other hand, the word *connect*, also in the 2K band, had an IF of 81.2 (Appendix C).

7. Discussion

The average test result for the 2K list (47%) suggests that pre-intermediate learners do not have receptive knowledge of half of the 1,000 words in the 2K frequency band, even after receiving seven weeks of instruction by mid-semester. The average test result for the 1K list (72%) suggests that 25% of the 1K list may be unknown to incoming pre-intermediate learners. If combined, the average learner entering pre-intermediate English is likely to have a gap of 750 high-frequency words (37.5% of the first 2,000 words). This finding clearly illustrates Browne's (1998) hypothesis that many pre-intermediate Japanese college students, while likely knowing more than 2,000 words, are not familiar with the most frequent 2,000 vocabulary items in English. Further, if the culture of test preparation leads to overreach in high school, university pressure to reach TOEFL targets can similarly result in neglecting the teaching and learning of frequent vocabulary in common course materials or in selecting inappropriate paths in individually-differentiated SRS systems. This creates the possibility of sustained overreach throughout the first two years in university English programs. If learners consistently try to read texts for which they do not have adequate coverage, they are likely to experience difficulty, lower test scores, and loss of motivation. Ensuring learners have maximal knowledge of high-frequency vocabulary and are developing fluent reading ability should be well-established practices, but these results suggest this may not be the case in Japanese high school programs.

The average false alarm rates of approximately 8% in the fall of 2013 and 9% in the fall of 2014 are similar to that found by Barrow et. al. (1999). Their learners claimed to know 1.26 of 15 nonwords, which would be 8.4%. This number is higher than most other researchers in Japan have found, however. Sevigny and Ramonda (2013) reported an average false alarm rate of 4.8%, which mirrors what Stubbe, Stewart, and Pritchard (2010) and Stubbe (2012) reported for low intermediate students (4-5%). There are a few factors that may have contributed to the higher false alarm rate in this study. First, in this study, 15 nonwords were added to each test of 110 words, which brought the total number of items on each test to 125. Nation (2008) recommended keeping tests to a maximum of 100 items. Sevigny and Ramonda (2013) only added five nonwords. Adding fewer nonwords might have helped reduce test fatigue. Additionally, one of the nonwords, *suddery*, resulted in a false alarm rate of about 36%. That is, about 36% of test-takers claimed to know it. This is over four times the average false alarm rate and appears to be an outlier among nonwords. A lower false alarm rate and correction factor, however, would not change the overall results significantly.

Nation (2008) stated that the most frequent missing vocabulary is best taught directly, and the results of a study like this can help teachers both select and recommend useful vocabulary lists and textbooks that fit with current needs of institutions, classrooms, and/or individuals. For example, the word *concern* had an IF of 19.2 and *indeed* had an IF of 25.0. These words are classified as 1K words on both the GSL and NGSL, yet only 19% of test-takers claimed to know the

word *concern*, and only 25% of learners claimed to know *indeed*. All of the following words had an IF below 50: *severe, literature, proper, plenty, explore, instrument, experiment, attempt, narrow, approve, extreme, thick, length, demand, row,* and *edge*. Arguably, these words would be extremely useful to add to a common course list—a list of words teachers use in teacher-fronted class activities. When teachers select common course materials, and thus, a one-size-fits-all approach to vocabulary, it is very important to have diagnostic information like that in Appendix C (the “Forgotten 400” unfamiliar words). Even with such information, common course materials often aim for the middle achievers, the average students, which is ineffective for high achievers in that they are exposed to material they already know. Likewise, for low achievers the opposite is true; they struggle to keep pace with shared material, and often fall further behind. It would be very difficult and time-consuming to overcome a 750-word deficit in vocabulary through common course materials or teacher-fronted activities alone. With this problem in mind, we have compiled a list of the “400 forgotten words” that represent the most frequent, missing words. This is a more realistic set for a common course list (Appendix C).

A further recommendation for action is to use individually-differentiated materials, which include students’ own personal notebooks of incidental vocabulary, flashcards (either computer adaptive or physical cards), and in the case of extensive reading, individually-chosen graded readers can be utilized for vocabulary learning. Since each student brings a different knowledge base to the classroom, these materials target individual gaps for students. Having individual learners review their own Yes/No test results for highly frequent vocabulary could help learners organize independent learning tasks. While individually differentiated materials cater to the individual needs of each student regardless of proficiency, these too have issues. First, access to these materials are often left up to students. Secondly, creation of these materials is time-consuming for students, and many of them do not know the best ways to create self-study materials. This is especially true for first-year university students, who often have been instructed in teacher-centered English classrooms. Being able to self-select vocabulary materials and study efficiently is a difficult task because students need to be given study materials and also need to be shown methods to study them as well. Considering these difficulties, a carefully-designed program of extensive reading with incidental vocabulary study, combined with spaced repetition software (SRS) flashcard study would provide a strong individual study track.

7.1 The “Forgotten 400” High Frequency Word List

In order to create this list, we combined the 1K and 2K results and narrowed the list down to 400 unfamiliar words. We selected words with item facilities of 80% or lower after subtracting the average false alarm rate of 8.33. Using the Web Vocabprofile (Cobb, 2010) with the added NGSL function, we only chose words from our list that also appear on the 1K and 2K NGSL lists and not words from other NGSL bands. In other words, the list of 400 words, we believe, contains relevant, highly frequent words that are unfamiliar to incoming Japanese university students in our program.

One outcome of this study is the understanding that we need to create a set of materials for pre-enrollment students, targeting specific gaps before they arrive on campus. Currently, our university offers an optional pre-enrollment English

program for students. This program is specifically targeted at students who have already passed the admission screening and typically place in the elementary or pre-intermediate level. Students enroll to improve their English skills before they get officially placed in the spring semester. A prerequisite to join the 2016 English pre-enrollment program is that all students must complete 50 study hours of a predetermined listening and reading program online. Upon completion, students are allowed to attend an 11-day English course, totalling 60 contact hours (40 *koma*). With new insights into the vocabulary deficiencies of incoming freshman students, one option is to redesign the pre-enrollment prerequisite around the “Forgotten 400” vocabulary. This, followed by a post-test at the beginning of the spring semester, could indicate whether these gaps have been addressed or if further study materials are required.

7.2 Limitations

Our study is not without some limitations. First, with a test size of 125 items, test fatigue might have become a factor. One improvement for the future would be to revise the test to create a shorter list of words—about 80 to 90 including nonwords that still represent the common core 1K and 2K adapted lists. Secondly, Yes/No tests are unable to measure the knowledge of homonymous words, or words with multiple meanings. When a student is tested on a highly contextual word such as *set*, it is unclear which definition the student is claiming to know. We would need to create contextualized Yes/No tests to determine which forms are unknown to our students. Another improvement could be to include other word forms and parts of speech within the test. However, this solution directly challenges the most prominent feature of Yes/No tests: they are an efficient way to test a large number of words. Adding more components to the current Yes/No tests will also only increase the test fatigue as well. Since we only performed tests that measured the receptive form-meaning connection of the 1K and 2K adapted GSL lists, we cannot claim to know the words students are able to produce in written or spoken forms. Furthermore, since our course objectives emphasize and test vocabulary knowledge in the form of speaking and writing assessments, it is a high priority to step back and rethink the exact number of words our students claim to know.

8. Conclusion

University educators need to consider the possibility of implementing vocabulary activities, assessments, and study programs to address the shift in vocabulary acquisition over time and contexts. There is always a need for university teachers to create and adapt coursework materials dynamically through and between semesters. This paper provides a strong argument for the reevaluation of the most frequent vocabulary known by Japanese university students and for creating a set of materials needed to fill in the gaps. This study has sought to demonstrate ways Yes/No tests can efficiently provide useful diagnostic data for relatively large word lists and organize the results into tailored lists that can inform both common course and individually differentiated modes of instruction. The gap discovered in the 2K range of high frequency vocabulary suggests that for many learners who enter demanding college English programs, this gap could easily develop into a chasm if not efficiently diagnosed and treated.

In some ways, vocabulary acquisition strategies have to be retaught to university English language students. Though the university can recommend *what* to study, learners also have to be given advice on *how* to study effectively. Although the teacher can play a major role in shaping the learners' vocabulary studies, learners have to be trained to reach a level of autonomy in their language learning. One part of effective autonomous learning involves introducing learners to SRS systems, as they offer a powerful tool for deliberate word study of both general and special lists. These computer adaptive systems, when added to carefully designed common course programs, promise to quickly help learners fill in their individual gaps. These systems, however, allow learners to choose paths of study, and thus educators will still need to guide learners to select the right paths, especially when SRS systems do not provide transparent data about vocabulary a student knows.

We have also shown the major gaps in the GSL vocabulary we believe exist for matriculating Japanese university students and have made recommendations on a course of action for teachers and administrators. Further research could begin to analyze the current vocabulary learning strategies Japanese university students employ and their effectiveness. Likewise, additional research could indicate how strategies used in traditional Japanese high schools differ from more autonomous learning mechanisms stressed in university English classrooms. University educators must be able to fill in the important vocabulary gaps in terms of both what to teach and how to guide students to become successful, autonomous, and critical language learners.

Our results only reflect a small portion of a student's total vocabulary, that is, passive, visual recognition of the form-meaning connection. It would also help to test passive, aural recognition of the sound-meaning connection. We cannot make strong claims about what productive vocabulary needs to be taught in the curriculum until we tease out which items students only know receptively and which ones they can produce. By implementing various types of vocabulary tests (passive and active), we can begin to triangulate receptive and productive vocabulary lists to maximize the vocabulary learning potential of English language learners at Japanese universities.

Acknowledgements

We would like to acknowledge Maiko Berger for her indispensable help in the JALT 2014 presentation of preliminary results.

References

- Barrow, J., Nakanishi, Y., & Ishino, H. (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System*, 27(2), 223-247.
- Browne, C. (1998). Japanese high school textbooks: How readable are they? *Temple University Japan Working Papers in Applied Linguistics*, 12, 1-13.
- Browne, C., Culligan, B., & Phillips, J. (2013). *New general service list*. Retrieved from

- <http://www.newgeneralservicelist.org>.
- Cihi, G. (2013). *Lexica's test score conversion methodology*. Retrieved from http://www.wordengine.jp/research/pdf/conversion_methodology2013.pdf
- Cobb, T. (2010). *Web Vocabprofile* [Computer Software] Accessed from <http://www.lex tutor.ca/vp/>, an adaptation of Heatley, Nation & Coxhead's (2002) Range.
- Cobb, T. (2010). *Plausible Nonword Lists from Paul Meara and Colleagues*. Accessed from http://www.lex tutor.ca/freq/lists_download/pnwords.html.
- Gyllstad, H., Vilkaite, L. & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL International Journal of Applied Linguistics*, 166, 276–303.
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403–430.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren, & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines*. Clevedon, UK: Multilingual Matters, Ltd.
- Mclean, S., Hogg, N., & Rush, T. (2014). Vocabulary size of Japanese university students: Preliminary results from JALT sponsored research. *The Language Teacher* 38(3), 34-37.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142–154.
- Mochida, A., & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. *Language Testing* 23(1), 73-98.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nation, I.S.P. (2008) *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle Cengage Learning.
- Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, 7(2), 105-125.
- Sevigny, P., & Ramonda, K. (2013). Vocabulary: What should we test? In N. Sonda & A. Krause (Eds.), *JALT2012 Conference Proceedings*. Tokyo: JALT.
- Shillaw, J. (1996). *The application of Rasch modelling to yes/no vocabulary tests*. Unpublished manuscript, University of Tsukuba, Ibaraki, Japan. Retrieved from <http://www.lognostics.co.uk/vlibrary/>
- Shillaw, J. (2009). Putting yes/no tests in context. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners: Papers and perspectives in honour of Paul Meara*, 13-24. Bristol, UK: Multilingual Matters.
- Stubbe, R., Stewart, J., & Pritchard, T. (2010). Examining the effects of pseudowords in yes/no vocabulary tests for low level learners. *Kyushu Sangyo University Language Education and Research Center Journal*, 5, 5-23.
- Stubbe, R. (2012). Do pseudoword false alarm rates and overestimation rates in yes/no vocabulary tests change with Japanese university students' English ability levels? *Language Testing*, Advance online publication. doi:10.1177/0265532211433033
- West, M. (1953). *A general service list of English words*. London: Longman.

Appendix A: Instructions

Methods: Instructions

This is a vocabulary test. Please indicate whether you know the word or not: "Yes" = "I know this word" or "No" = "I don't know." By "knowing" a word, we mean that you are able to recognize its basic meaning.

これは語彙知識を判断するテストです。それぞれの単語を知っているかどうか、該当する選択肢を選んでください。単語を「知っている」ということは、その単語の基本的な意味が分かるということです。

Appendix B: False Alarm Rates

Nonwords & False Alarm Rates (FAR)			
Nonwords	Fall 2013 FAR	Fall 2014 FAR	Total AVG FAR
moffat	2.94	0.34	1.64
cantileen	1.78	2.99	2.385
opie	2.4	3.38	2.89
adair	2.98	3.33	3.155
hubbard	4.77	2.71	3.74
runster	4.43	X	4.43
whitrow	3.57	5.31	4.44
kiley	X	5.6	5.6
duffin	4.73	6.48	5.605
cambule	5.98	6.89	6.435
ridout	5.23	10.56	7.895
glandle	5.88	9.98	7.93
twose	7.22	10.84	9.03
bance	12.84	15.26	14.05
ralling	17.57	18.65	18.11
suddery	33.88	37.85	35.865
Average	7.75	9.34	8.33

Appendix C: The Forgotten 400

Forgotten 400 Unfamiliar Words: 1K NGSL (133 items) and 2K NGSL (267 items)					
Word	IF (Adjusted)	Word	IF (Adjusted)	Word	IF (Adjusted)
accuse	12.2	broad	59.2	weak	73.7
arrest	12.2	mention	59.2	mouse	73.8
debt	12.8	suffer	59.2	lay	73.8
discipline	13.7	union	59.2	temperature	73.9
folk	15.4	within	59.2	matter	74.0
severe	15.4	minister	59.3	contain	74.2
reputation	17.3	property	59.3	content	74.2
dozen	17.4	struggle	59.3	fill	74.2
qualify	17.4	oppose	59.3	nor	74.2
column	18.4	grey	60.1	opportunity	74.2
concern	19.2	stuff	60.1	supply	74.2
gene	22.5	weak	60.1	deliver	74.6
loan	23.3	journey	60.3	tie	74.6
indeed	25.0	observe	60.9	tools	74.6
literature	25.0	progress	60.9	detail	74.6
delight	25.0	spread	60.9	disappoint	74.6
victim	25.0	straight	60.9	parent	74.6
bind	25.8	tend	60.9	slide	74.6
reward	27.6	calculate	61.1	swing	74.6
meanwhile	27.8	effort	61.7	remain	75.0
council	29.2	equal	61.7	storm	75.0
wage	29.5	permit	61.7	lead	75.5
breast	29.9	prevent	61.7	per	75.5
rural	30.2	regard	61.7	rather	75.5
proper	30.2	sight	61.7	separate	75.5
crisis	31.2	forth	62.0	throw	75.5

defeat	31.7	represent	62.0	wide	75.5
blame	32.7	absence	62.3	balance	75.9
chain	32.7	calm	62.3	knock	75.9
crop	32.7	comfort	62.3	shock	75.9
threat	33.4	hall	62.4	complain	76.1
witness	33.4	typical	62.4	gun	76.1
aside	33.8	attract	62.8	suit	76.1
profit	34.6	competition	62.8	wine	76.1
arise	35.0	patient	62.8	although	76.3
provision	35.0	dead	63.1	instead	76.3
guard	35.3	hill	63.1	brown	76.3
plenty	35.3	tour	63.5	neck	76.3
urban	36.1	reflect	63.9	passenger	76.3
explore	36.4	extend	64.2	recommend	76.3
instrument	36.4	strength	64.2	ancient	76.7
wheel	36.4	above	64.7	appoint	76.7
distinguish	36.7	divide	64.7	bill	76.7
enemy	36.7	latter	64.7	due	76.7
experiment	36.7	vote	64.7	political	76.7
determine	37.4	sudden	64.9	purpose	76.7
attempt	37.7	religion	65.0	raise	76.7
suspect	37.9	bomb	65.0	stock	76.7
narrow	38.4	essential	65.4	extra	77.0
cent	40.3	confidence	66.0	frame	77.0
deserve	40.3	moreover	66.0	habit	77.0
launch	40.3	defense	66.1	hire	77.0
declare	40.3	command	66.7	bottom	77.1
efficient	40.3	flow	66.7	climb	77.1
honor	40.4	judge	66.7	replace	77.1
firm	40.6	manufacture	66.7	shout	77.1

practical	40.6	prison	66.7	bridge	77.4
approve	41.7	soldier	66.7	lake	77.4
extreme	41.7	tough	66.7	nature	77.4
faith	41.7	border	67.3	compare	77.4
joint	41.7	educate	67.3	conversation	77.4
grateful	42.9	emotion	67.3	guide	77.4
path	42.9	warn	67.3	opposition	77.4
absolutely	43.0	encourage	67.3	pair	77.4
pause	43.0	responsible	67.3	risk	77.4
weapon	43.0	court	67.4	sheet	77.4
affair	43.1	pound	67.4	thin	77.4
settle	44.2	passage	68.0	argue	77.8
advertise	44.5	pattern	68.2	band	77.8
thick	44.5	amount	68.4	rare	77.8
delay	44.6	except	68.4	slip	77.8
murder	44.6	expense	68.4	act	78.2
fault	45.4	surface	68.4	advantage	78.2
scale	45.4	refuse	68.6	blood	78.2
entire	45.5	proud	68.6	daughter	78.2
former	45.7	rush	68.6	either	78.2
examining	46.0	force	68.8	exist	78.2
accord	46.7	gather	68.8	general	78.2
fellow	46.7	moral	68.8	metal	78.2
length	46.7	plain	68.8	social	78.2
upon	46.7	recent	68.8	while	78.2
immediate	47.0	bone	68.8	beauty	78.4
tone	47.0	adopt	69.2	blow	78.4
upper	47.0	beyond	69.2	mouth	78.4
permanent	47.3	claim	69.2	organize	78.4
lean	47.6	current	69.2	succeed	78.4

tip	47.6	doubt	69.2	complicate	78.5
defend	47.8	influence	69.2	kick	78.5
agent	48.1	mister	69.2	load	78.5
jointed	48.4	silence	69.2	stretch	78.5
occasion	48.4	strike	69.2	birth	78.9
prove	48.4	climate	69.5	block	78.9
demand	48.5	impression	69.5	engine	78.9
mass	48.5	refer	69.5	import	78.9
otherwise	48.8	yard	69.5	sharp	78.9
edge	49.5	neighbor	70.1	steal	78.9
row	49.6	relative	70.1	board	78.9
ideal	50.5	society	70.1	capital	78.9
breathe	50.7	therefore	70.1	admit	79.2
compete	50.7	gap	70.6	destroy	79.2
intend	50.7	net	70.6	employ	79.2
district	51.2	treat	70.6	fact	79.2
ordinary	51.2	retire	71.1	lady	79.2
secretary	51.2	satisfy	71.1	law	79.2
afford	51.7	degree	71.2	middle	79.2
desire	51.7	duty	71.2	recognize	79.2
earn	51.7	particular	71.2	scene	79.2
remark	51.7	term	71.2	surround	79.2
sort	51.7	toward	71.2	coat	79.5
split	51.7	decrease	71.7	repeat	79.5
yield	51.7	figure	71.7	sample	79.5
youth	51.7	flat	71.7	unclear	79.5
disease	52.2	gain	71.7	avoid	79.9
associate	53.2	measure	71.7	beat	79.9
criminal	53.2	nation	71.7	breath	79.9
plate	53.2	prepare	71.7	host	79.9

seed	53.2	pressure	71.7	perform	79.9
coast	53.3	shall	71.7	double	80.3
army	54.2	signal	71.7	especially	80.3
thus	54.6	spell	71.7	pack	80.3
reference	54.9	suppose	71.7	effect	80.3
article	55.0	tear	71.7	increase	80.3
committee	55.0	waste	71.7	round	80.3
elect	55.0	behavior	72.2	variety	80.3
industry	55.0	stocking	72.2	whether	80.3
justice	55.0	root	72.3	abroad	80.6
capture	56.1	burn	72.8	female	80.6
among	56.7	fair	72.8	shut	80.6
certain	56.7	manner	72.8	attend	81.2
favour	57.4	perhaps	72.8	connect	81.2
ought	57.4	press	72.8		
combine	57.5	serve	72.8		
shoe	57.6	unless	72.8		
arrange	58.4	hole	73.3		
preserve	58.4	branch	73.7		