# Vocabulary: What Should We Test?

Paul Sevigny
Ritsumeikan Asia Pacific University
Kris Ramonda
Kwansei Gakuin University

**Reference Data:**
Sevigny, P., & Ramonda, K. (2013). Vocabulary: What should we test? In N. Sonda & A. Krause (Eds.), *JALT2012 Conference Proceedings*. Tokyo: JALT.

Diagnostic Yes/No tests are a recommended and much researched assessment tool (Read, 2007; Nation, 2008), yet there is little research into how to apply them to address the mismatch between pre-existing course vocabulary lists from commercial textbooks for a particular level and learners' actual vocabulary knowledge. This study looked at a vocabulary battery of 240 words adopted with a textbook for a pre-intermediate level English course at a Japanese university. During the 1st week of instruction, a Yes/No test including nonwords (pseudo-words) was administered in three forms with 85 items each. Approximately 100 students took each form. On the average, test takers claimed they knew 75% of the items on the list. A low false alarm rate supports Shillaw's (1996) findings that the use of nonwords could be lessened significantly in the Japanese context.

Yes/No形式の語彙診断テストは、学力診断用に奨励されかつ研究がされているツールである（Read, 2007; Nation, 2008）が、特定のレベルおよび学習者用に使用される市販教科書と、既に存在している語彙リストとの齟齬を分類する方法に関する研究は、ほとんど存在しない。当研究では、日本の大学における準中級英語科目で使用されている教科書から抽出した240語の語彙群について報告するものである。講義開始から1週間の間に、無意味語を含む85語ずつ3種類のYes/No形式の診断テストが行われた。各テストをおよそ100名ずつの学生が受けた。平均して、受験者はリストの75%の単語を熟知語であると判断した。無意味語を「知っている」と回答した割合が低かったことから、無意味語の使用は日本では大幅に減らすことができるというShillaw（1996）の結果を肯定する結果となった。

The initial impetus for this study came from a top-down English program evaluation and the development of curricular research teams. Teacher researchers were concerned that preexisting vocabulary achievement tests were being used for grading purposes without determining the learners' knowledge of the items being tested at the beginning of the semester. The problem was the need for content validation with respect to vocabulary related goals and testing. Assessing learners' knowledge with a Yes/No test at the beginning of the semester was the first step in evaluating the common vocabulary list and achievement tests in a program with large reading and vocabulary classes at a Japanese university.

## Literature Review

Yes/No tests traditionally present a word without context and ask the participant to indicate whether the word is known or not. In this report *checklist* will be used interchangeably with

Yes/No test, as the difference only refers to whether the learner circles, checks, or clicks to indicate whether an item is known or not. Nation (2008) and Read (2007) recommended the use of Yes/No vocabulary tests (in addition to the use of Vocabulary Levels Tests [VLTs]) for placement purposes. While the VLT provides general placement information with regard to vocabulary frequency bands, Read suggested that Yes/No tests can assess vocabulary size relative to specific lists, and thus may be used by programs to develop their own assessment tools. Additionally, because of its simplicity, the Yes/No format is "informative and cost effective" (Read, 2007, p. 113).

Shillaw (2009) presented a careful overview of Yes/No tests and the many efforts to establish the reliability and validity of the tests. Zimmerman, Broder, Shaughnessy, and Underwood (1977) measured the vocabulary size of native English speakers and reported validity of a word recognition test based on correlation with verbal scores on the Scholastic Aptitude Test. Anderson and Freebody (1983) studied the vocabulary size of 5th grade native English speakers and reported that Yes/No test results were more reliable than multiple-choice tests for measuring vocabulary size. This led to the first research into Yes/No tests for determining the size of L2 learners' vocabulary. The rationale was that the simplicity and efficiency of the Yes/No format would allow for sampling the large number of items necessary for estimating vocabulary size. Meara and Buxton (1987) compared a Yes/No test with a multiple-choice test in predicting nonnative speakers' grades on the First Certificate in English examination and reported that only the Yes/No scores had a significant correlation with grades. By the turn of the century, these tests had become established measures of vocabulary size for L2 learners (Nation, 2001; Read, 2000). In support of Yes/No tests, Cameron (2002) stated, "Eventually, after sufficient contextualized encounters, a word will be recognized when it is met in context or in isolation . . . it does not seem unreasonable to test to see how much vocabulary can be recognized without extended linguistic or textual contexts" (p. 151).

More recently, Mochida and Harrington (2006) completed an in-depth review of the Yes/No test as a tool for testing receptive vocabulary. They reviewed a number of studies that attempted to correlate results from Yes/No tests with other test forms—multiple-choice and translation tests. They stated, "The results show that the Yes/No test is a reliable measure of the kind of vocabulary knowledge measured by the VLT and, presumably, similar multiple-choice tests" (p. 91). Finally, they concluded that the Yes/No test has compelling practical advantages warranting further attention from L2 testers and teachers, such as incorporating their use in word recognition tasks.

The use of nonwords started as an attempt to validate test takers' judgments. The nonwords follow the phonetic rules of English and provide a window into determining whether students are honestly stating their familiarity or unfamiliarity with vocabulary items. Nonwords are typically created by changing one or two letters in real words; for example, *foggy* becomes *wuggy*. A second method of creating nonwords, called pseudo-derivation, uses unconventional base + affix combinations, for example, *adjustation* (Shillaw, 2009). A false alarm is an instance when a learner reports knowing a nonword. Read (2007) recommended using nonwords to correct the total score for each learner on a Yes/No test by simply taking the number of Yes/No responses to real words minus the number of Yes responses to nonwords and finding the resulting vocabulary size.

Calculating average false alarm rates for populations allows for identifying populations that are generally overconfident. In a study done in Japan (Barrow, Nakanishi, & Ishino, 1999), participants reported knowing an average of 1.26 nonwords on tests with 15 nonwords, yielding a false alarm rate of 8.4%. Milton (2009) provided averages from studies outside Japan with much higher false alarm rates. Shillaw (1996) found Japanese learners

to be very conservative, almost never falsely claiming knowledge of nonwords. Stubbe, Stewart, and Pritchard (2010) and Stubbe (2012) reported false alarm rates for low-intermediate Japanese university students of 4-5%.

Perhaps the most compelling of studies attempting to investigate the reliability and validity of Yes/No tests with the use of nonwords was Shillaw's (1996) use of Rasch scaling techniques to examine three of Meara's (1992) Yes/No tests that were each comprised of 40 real words and 20 nonwords. These tests were administered to seven classes of Japanese university students. In a rather complex Rasch analysis comparing the results of two tests taken by the same groups of students, Shillaw reported higher correlations when comparing scores of real words only versus the scores which included all words (real words and nonwords). Shillaw pointed out the marginal value of nonwords in contributing to test variance; he concluded that on these assessments and for these learners, the presence of nonwords had little effect on their test performance.

## Context of the Study

The context for the current study was the pre-intermediate English level of a large EFL program at a Japanese university. The courses were divided by skills into a two-credit reading and vocabulary course and a four-credit listening, speaking, writing, and grammar course. This research was situated in the pre-intermediate English reading and vocabulary course. At the time of this study, learners were placed using the paper-based TOEFL test without a diagnostic VLT. The course utilized a commercial textbook and aimed to increase learners' receptive vocabulary knowledge for readings in the textbook.

The English Program followed a 5-year curriculum cycle, implementing an all-new curriculum every 5 years. Prior to the beginning of the new curriculum cycle in 2011, textbooks were adopted as a base for each course. In the pre-intermediate level, *Interactions Access: Reading* (Hartmann & Mentel, 2007) was selected. The first seven chapters and corresponding 240 items in chapter word banks were adopted as the base curriculum and common course vocabulary list. The items were taught and tested in the first two semesters of the curriculum cycle. Various computer-based vocabulary activities were constructed based upon the new common list for individual student practice. Regular, summative vocabulary assessments, accounting for 30% of learners' grades, consisted of multiple-choice and matching items and tested receptive, form-meaning connections (the ability to recognize a word and recall its meaning).

After two semesters in the new curriculum cycle, students' average score on vocabulary quizzes and tests was above 95%. Although high scores are encouraging to all stakeholders, teachers began to voice concerns that the high grades might be a result of learners already knowing the vocabulary. The current study began with the desire to ensure that learners have a worthwhile learning experience.

## Research Issues

The goal of this research was to determine how much of the established course vocabulary list learners believed they already knew at the beginning of the course. For the purpose of this study, the extent of that knowledge was considered for the group as a whole and not for individuals. That is, the measure of interest was the percentage of learners who reported they knew a particular item. A further goal was to determine the corpus-based frequency of each item.

The following research questions were investigated:

1. How much of the common course vocabulary list is already known by most of the learners?

2. How does the learners' familiarity with each word (item facility) relate to established corpus-based frequency-band data?

## Method

### Participants

The participants in this research were 300 university learners of pre-intermediate English (TOEFL scores 400-437). Of these learners, 89% were Japanese and the remaining 11% were learners from China and Korea who were already fluent in Japanese. All participants were 1st-year university students, most having placed directly into pre-intermediate English, but some continuing from a previous semester in elementary English.

### Procedure

#### Frequency Information

The first step was to record item frequencies for all 240 words (see Appendix, column *f*) from Web Vocabprofile (Cobb, 2006), an adaptation of Heatley and Nation's (1994) Range. The profiler provides lemmatized word frequency information: K1, K2, AWL, and Off List; that is, the same frequency is assigned to all members of one lemma, or headword. For example, the word *problematic* is a member of the word family *problem*, which falls in the K1, or most frequent one-thousand word families. Following are item frequency categories:

- K1—word from first thousand most frequent word families,
- K2—word from second thousand most frequent word families,
- AWL—word from Academic Word List, and
- Off List—word not included in K1, K2, and AWL.

#### Yes/No Test

The 240 items of the common course list were organized by frequency, alphabetized, and then divided into three groups to make three test forms. The intention was to create three sets of words that presented variety as to frequency, spelling, meaning, and word length, rather than having alphabetically ordered segments or chapter themes grouped together. Similar-sounding items were intentionally separated when possible. The rationale for having three separate test forms was out of a concern that test taker fatigue could impact participant responses. Nation (2008) suggested using from 50 to 100 items in such an assessment. In the current study, 80 items were used in each test. The same five nonwords were added to control for overconfidence, yielding a total of 85 words on each Yes/No test. The downloadable application Wuggy 0.2.0b3, available from the Center for Reading Research at Ghent University, was used to create the following nonwords of similar length: *wuggy*, *ecution*, *pregime*, *mengel* and *runster*. The word lists from the three tests (A, B, and C) are in the Appendix.

After their level was determined via the school placement test, learners were randomly assigned to classes. Learners for this study came from six teachers' pre-intermediate English classes (two teachers' groups completed each test form). There were three large classes (60+ students) and three slightly smaller classes (40+ students). Large and small classes were paired to form three groups of approximately equal size, each of which received one test: Test A, Test B, or Test C. The three tests were administered on the second day of instruction during the spring semester of 2012. Each test was administered to two classes. See Figure 1 for instructions and examples.

### Figure 1. Instructions and Examples From the Yes/No Vocabulary Tests

Instructions:
This is a vocabulary test. Please indicate whether you know the word or not, "I know this word" or "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。

Examples

1. travel
   a) I know this word.
   b) I don't know.

2. wuggy
   a) I know this word.
   b) I don't know.

## *Scoring*

The test was administered using a content management system called Blackboard 6.2. The students indicated their choice by clicking a radio button. Blackboard 6.2 yields score reports for individual learners, group averages, and individual test items, including the percentage of learners answering each item correctly. Three scoring procedures were used in this study:

1. *average test score:* the average percent of the real words that the learners claimed to know,
2. *item facility:* the percentage of learners who claimed to know that item, and
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.

## Results

Blackboard 6.2 does not automatically save learners' answers during a test. Those who do not use the "save" or "save all" function receive a zero. If a test taker does not save any answers, it is evident as all item responses appear as "not answered." There were 34 participants who did not save their answers and thus were removed from the study, resulting in a sample size of 300.

## *Average Test Scores*

Average test scores are reported in Table 1. The total number of words known minus the number of nonwords (incorrectly selected) yields vocabulary size (Read, 2007). The results showed that almost all learners correctly rejected all five nonwords. The adjusted average vocabulary size using Read's formula would yield about 59 out of 80. In other words, on the second day of class, learners claimed to know about 75% of all words on the common course list.

### Table 1. Average Test Scores ($N$ = 300)

| Test | Average number known words ($k$ = 80) | Average number non-words correctly rejected ($k$ = 5) |
|---|---|---|
| A ($n$ = 111) | 55 (69%) | 4.7 (94%) |
| B ($n$ = 101) | 61 (76%) | 4.8 (97%) |
| C ($n$ = 88) | 64 (80%) | 4.8 (96%) |
| Total | 60 (75%) | 4.8 (96%) |

*Note.* $k$ = number of words on test

### Item Facility Results

The percentage of learners reporting to know each test item is reported in the Appendix. In classical testing theory, this measure is known as the item facility and is calculated by taking the number of students who reported to know a word, divided by the total number of responses. They are listed in descending order in percentage form, with the items that 100% of all learners reported knowing at the top. The frequency band from Web Vocabprofile for each item is reported to the right of each item facility result.

## Discussion

### False Alarm Results

The false alarm average of 4.8% reveals that the learners in this study were reasonably conservative when deciding whether an item was known or not. As we added just 5 nonwords to each test containing 80 real words, we presented learners with 94% real words on the test. Meara (1992) presented 40 real words out of a total of 60 words, or a proportion of 67% real words. Due to time constraints and concern about test-taker fatigue, we opted for a smaller pool of nonwords. More importantly, in the light of the commonly low false alarm rates reported by other researchers in Japan (Shillaw, 1996; Stubbe, 2012; Stubbe et al., 2010), decreasing the proportion of nonwords in the Japanese context seemed reasonable.

### Item Frequencies

While there are clear visual correlations between the K1 words and the most familiar items at the top of the Appendix, any obvious pattern obscures with decreasing item facility. Applying more sophisticated statistical analyses to the results in the Appendix could provide more nuanced understanding of what makes some items easier than others to acquire. For example, item facility scores give an indication of learners' familiarity with collocations, for which frequency data is not readily available. In some instances where there is disparity between the item facility and item frequency, there may be a need to reclassify item frequencies. For example, *computer* is listed as AWL, but now is likely to be in the most frequent one thousand words. *Email* and *online* are OL, but the frequency of these items has also increased dramatically. Meara (2010), in the preface to the second edition of his *EFL Vocabulary Tests*, cited the effect of digital communications on word frequency:

> *Text*, for example, was a relatively infrequent word in 1992, largely confined to a couple of very specific genres. Thanks to mobile phones, *text* must be one of the most frequent words to occur in everyday spoken English in 2010. (p. 3)

### Knowing a Word Receptively

One limitation of the Yes/No test for receptive knowledge, as it was used in the current study, might be the provision of the decontextualized, written word alone. There is more to knowing a word receptively than just seeing it—for example, knowing a word by hearing it. Another possible way to construct this test could have been to include both the written word and an audio recording of the same word. Learners could have worn headphones and clicked a button to play the audio pronunciation as they looked at the written form of the word. Providing both audio and visual channels could impact the results and might serve to identify words that have already been partially learnt receptively. In the current study, the primary interest was in measuring participant knowledge of the form-meaning of the target items, since the course vocabulary tests were designed in a similar fashion. For these reasons, only the written form-meaning was presented.

A second way to alter the form of a Yes/No test would be to provide context around the given vocabulary item, that is, to provide the word in a clear sentence and have the learner indicate whether the underlined word is known or unknown. The added dimension of sentence level context in Yes/No tests was investigated by Shillaw (2009). He compared the results of two Yes/No tests, one with context provided (with instructions that encouraged test takers to use it) and one with no context for the lexical items. The results showed a statistically significant and higher affirmation rate for the Yes/No test with context provided. If sentence level context can trigger the learner to recognize an item, then the item may already be partially acquired receptively.

The decontextualized Yes/No test given in this study provides less receptive context than either of the two alternatives above. This lack of added context presents the least likely reading scenario for learners in a natural context; thus, it might be logical to infer that the affirmation rate on the Yes/No test in this study, with the addition of aural or sentence context, would increase if listening or context were added. When encountering words in authentic contexts, a number of other linguistic features such as aural, visual, and syntactic cues could aid in accessing partially learnt receptive lexical items. On one hand, there is the possibility that with added context the learner might not actually know the word, but could infer the meaning from its lexical environment. Furthermore, a context-rich environment could place an additional burden on working memory, especially if other words were unknown to the learner, or presented in cognitively difficult-to-process grammatical structures. This could potentially distract the learner from the target item. Moreover, adding these other dimensions also takes more time, both in creating and taking the test, which would diminish some of the simplicity and efficiency that makes the Yes/No test so valuable. Nonetheless, studying how added context in Yes/No tests affects the reliability of the results deserves more attention.

## Conclusion

When curriculum or learner populations (or both) are in flux, it is inevitable for teachers and administrators at some point to ask the question posed in the title of this paper—Vocabulary: What should we test?

Language programs should include tests of vocabulary levels. However, program administrators need to test lexical knowledge that corresponds to both the needs of the learners and the levels of courses in which learners will be placed. Thus, they need something different than a norm-referenced test like the VLT. Diagnostic Yes/No tests, as employed in this study, could provide administrators with valuable item facility data for creating custom placement tests. Diagnostic Yes/No tests can also provide important data about the level of mastery a cohort of learners has in relation to a common course list so that teachers can move specific vocabulary items from receptive to productive-mode tasks and assessments. The results of Yes/No tests can also move a course that has adopted a specific textbook towards the process of *adapting* the use of that textbook's vocabulary lists.

The mismatch between predetermined vocabulary lists in commercial textbooks for a particular level and students' actual vocabulary knowledge in corresponding levels can lead to inefficient allocation of teacher and time resources. This study provides one possible solution to address this mismatch. The Yes/No test is quick and easy to administer and allows for agility and flexibility in tailoring vocabulary items to a specific student population. However, this type of assessment is not without limitations. There can be reliability issues due to overconfidence or misinterpretation of what it means to know a word on the part of the test taker. This can be controlled to some extent by including nonwords in the assessment and by providing clear examples of what knowing a word means in the test instructions.

## Bio Data

**Paul Sevigny** is a Lecturer at the Center for Language Education, Ritsumeikan Asia Pacific University in Beppu, Japan. His areas of research interest are extensive reading, vocabulary, and the diffusion of best practices.

**Kris Ramonda** is an Associate Lecturer of English at Kwansei Gakuin University in Japan. His research interests include vocabulary acquisition, extensive reading, and metaphor in language.

## References

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In Hutson, B. A. (Ed.), *Advances in reading/language research*, 2 (pp. 231-256). Greenwich, CT: JAI Press.

Barrow, J., Nakanishi Y., & Ishino, H. (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System*, *27*, 223-247. doi:10.1016/S0346-251X(99)00018-4

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research*, *6*, 145-173. doi:10.1191/1362168802lr103oa

Cobb, T. (2006). *Web Vocabprofile* [Computer program]. Accessed from http://www.lextutor.ca/vp/

Hartmann, P., & Mentel, J. (2007). *Interactions access: Reading*. Singapore: McGraw-Hill.

Heatley, A., & Nation, P. (1994). *Range* [Computer program]. Victoria University of Wellington, New Zealand. Available from http://www.vuw.ac.nz/lals/

Meara, P. (1992). *EFL vocabulary level tests*. Swansea, UK: University of Wales, Swansea.

Meara, P. (2010). *EFL vocabulary level tests* (2nd ed.). Swansea, UK: University of Wales, Swansea.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*, 142-154. doi:10.1177/026553228700400202

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, UK: Multilingual Matters.

Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, *23*, 73-98.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle Cengage Learning.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, *7*(2), 105-125.

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* (pp. 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

Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. (1977). A recognition test of vocabulary using signal-detection measures and some correlates of word and nonword recognition. *Intelligence 1*(1), 5-13.

# Appendix

## *Pre-Intermediate Course Vocabulary Yes/No Test Results*

### *Results: Yes/No % and Item Frequencies*

| Test A $n = 111$ | K% | *f* | Test B $n = 101$ | K% | *f* | Test C $n = 88$ | K% | *f* |
|---|---|---|---|---|---|---|---|---|
| easy | 100.0 | K1 | address (n) | 100.0 | K1 | alone | 100.0 | K1 |
| move (vb) | 100.0 | K1 | building (n) | 100.0 | K1 | child (children) | 100.0 | K1 |
| people | 100.0 | K1 | carry | 100.0 | K1 | cities (city) | 100.0 | K1 |
| too (adv) | 100.0 | K1 | different | 100.0 | K1 | country (countries) | 100.0 | K1 |
| computer | 99.1 | AWL | in | 100.0 | K1 | live (vb) | 100.0 | K1 |
| fun | 98.2 | K2 | mean (vb) | 100.0 | K1 | power (n) | 100.0 | K1 |
| however | 98.2 | K1 | money | 100.0 | K1 | problem | 100.0 | K1 |
| life | 98.2 | K1 | team | 100.0 | AWL | travel (vb) | 100.0 | K1 |
| small | 98.2 | K1 | women | 100.0 | K1 | busy (adj) | 98.9 | K2 |
| wonderful | 98.2 | K1 | work (vb) | 100.0 | K1 | large | 98.9 | K1 |
| active | 97.3 | K1 | choose | 99.0 | K1 | monster | 98.9 | OL |
| store (n) | 97.3 | K1 | drive (vb) | 99.0 | K1 | put (vb) | 98.9 | K1 |
| married | 96.5 | K1 | energy | 99.0 | AWL | information | 98.9 | K2 |
| teach (taught) | 96.5 | K1 | from | 99.0 | K1 | aunt | 97.8 | K2 |
| introduce | 96.4 | K1 | second (adj) | 99.0 | K1 | grow | 97.8 | K1 |
| uncle | 95.6 | K2 | famous | 98.0 | K1 | plant (vb) | 97.8 | K1 |
| across (adv) | 94.7 | K1 | generation | 98.0 | AWL | population | 97.7 | K1 |
| afraid | 94.7 | K2 | on | 98.0 | K1 | revolution | 97.7 | AWL |
| bag (n) | 94.7 | K2 | volunteer (n) | 98.0 | AWL | volleyball | 97.7 | OL |
| similar | 94.7 | AWL | public (adj) | 96.0 | K1 | email (vb) | 96.6 | OL |
| customer | 93.8 | K2 | take care of | 96.0 | K1 | in front of | 96.6 | K1 |
| search (vb) | 93.8 | K2 | neighbor | 94.1 | K1 | outside (adj) | 96.6 | K1 |
| draw (drew) | 92.0 | K1 | position (n) | 94.1 | K1 | sell (vb) | 96.6 | K1 |
| gym | 91.2 | OL | daily | 93.1 | K1 | stage (n) | 96.6 | K1 |
| research (n) | 91.2 | AWL | percent | 93.1 | AWL | feelings | 95.5 | K1 |
| close (adj) | 90.3 | K1 | huge | 91.1 | OL | online | 94.3 | OL |
| quit | 89.3 | OL | private (adj) | 91.1 | K1 | scientist | 94.3 | K1 |
| corner (n) | 88.5 | K2 | street children | 91.1 | K1 | symbol | 93.3 | AWL |
| volume | 87.6 | AWL | terrible | 91.1 | K2 | wonder (vb) | 93.3 | K1 |

| Test A $n = 111$ | K% | *f* | Test B $n = 101$ | K% | *f* | Test C $n = 88$ | K% | *f* |
|---|---|---|---|---|---|---|---|---|
| rent (n) | 86.7 | K2 | transport (n) | 91.1 | AWL | double (vb) | 93.2 | K2 |
| environment | 85.0 | AWL | average (n) | 90.1 | K1 | homeless (adj) | 93.2 | K1 |
| wedding | 82.3 | OL | either | 90.0 | K1 | neighborhood | 92.0 | K1 |
| familiar | 81.4 | K1 | lonely | 88.1 | K2 | dirty (adj) | 91.1 | K2 |
| mall | 77.9 | OL | product | 88.1 | K1 | drugs | 91.0 | OL |
| fix (vb) | 76.8 | K1 | journalist | 87.1 | OL | entertainment | 90.9 | K2 |
| unimportant | 76.1 | K1 | repair (vb) | 87.1 | K2 | according to (prep) | 88.6 | K1 |
| childhood | 74.3 | K1 | vision | 87.1 | AWL | full time | 88.6 | |
| available | 73.5 | AWL | habit | 86.1 | K2 | realize | 88.6 | K1 |
| olive tree | 72.6 | | single-parent family | 86.1 | | category | 87.6 | AWL |
| traditional family | 70.8 | | unfair | 86.1 | K1 | contain | 87.5 | K1 |
| cousin | 69.9 | K2 | benefit (n) | 85.1 | AWL | deliver | 87.5 | K2 |
| retire | 69.9 | K2 | equal (adj) | 85.0 | K1 | publish | 87.5 | AWL |
| improvement | 69.0 | K2 | release (vb) | 84.2 | AWL | traditional | 87.5 | AWL |
| influence (n) | 69.0 | K1 | purpose (n) | 83.2 | K1 | awake | 85.2 | K2 |
| suggestion | 69.0 | K1 | apologize | 82.2 | K2 | prepare | 85.1 | K1 |
| calculate | 68.1 | K2 | financial (adj) | 82.2 | AWL | crowded (adj) | 84.3 | K1 |
| demonstrate | 68.1 | AWL | charity (n) | 81.2 | OL | conversation | 83.9 | K2 |
| psychologist | 66.4 | AWL | inform | 81.2 | K2 | version | 83.9 | AWL |
| emotions | 65.5 | OL | focus on | 79.2 | AWL | basics | 83.1 | OL |
| virtual shopping mall | 65.5 | | academic (adj) | 76.2 | AWL | teenager | 83.0 | OL |
| contrast (n) | 64.6 | AWL | crime | 74.5 | K2 | site (n) | 82.8 | AWL |
| tend to | 62.8 | K2 | uninteresting | 73.3 | OL | responsibility | 81.8 | K2 |
| branch (n) | 61.9 | K1 | AIDS (n) | 73.0 | AWL | occur | 81.6 | AWL |
| behavior | 61.1 | K2 | orders (n) | 72.3 | K1 | marriage | 79.5 | K1 |
| logic | 61.1 | AWL | tough | 72.3 | K2 | gather | 77.3 | K1 |
| gender | 60.2 | AWL | barbecue (n) | 71.3 | OL | replace | 77.0 | K2 |
| desires (n) | 59.3 | K1 | great-grandparents | 69.3 | | application (app) | 73.9 | K1 |
| nuclear families | 58.4 | | survey (n) | 67.3 | AWL | wealth | 72.7 | K1 |
| predict | 58.0 | AWL | garage (n) | 66.3 | K2 | unfamiliar | 69.3 | K1 |
| portable | 57.5 | OL | argue | 65.3 | K2 | relatives | 69.0 | K1 |
| homelessness | 56.6 | OL | mammal | 65.3 | OL | complicated (adj) | 68.2 | K2 |
| reward (n) | 52.2 | K2 | politics | 65.3 | K1 | megacity | 65.5 | OL |

| Test A $n = 111$ | K% | *f* | Test B $n = 101$ | K% | *f* | Test C $n = 88$ | K% | *f* |
|---|---|---|---|---|---|---|---|---|
| symbolize | 50.4 | AWL | cost of living | 63.4 | | make sense | 64.8 | |
| point out | 46.0 | K1 | satisfaction | 61.4 | K2 | profit (n) | 63.2 | K1 |
| anxious | 45.1 | K2 | emotional | 56.4 | OL | take responsibility | 62.8 | |
| divorce (n) | 44.2 | OL | competition | 55.4 | K2 | struggle (vb) | 59.8 | K1 |
| evidence (n) | 44.2 | AWL | resident | 52.5 | AWL | illegal (adj) | 57.5 | AWL |
| feminist | 44.2 | OL | socialize | 49.5 | OL | adulthood | 57.0 | AWL |
| appropriate (adj) | 40.2 | AWL | annual (adj) | 41.6 | AWL | adulthood | 56.8 | AWL |
| mixture | 37.2 | K2 | conflict (n) | 39.6 | AWL | emotionally | 56.8 | OL |
| purchases (n) | 37.2 | AWL | donate | 37.6 | OL | household | 53.5 | OL |
| embarrassing (adj) | 36.8 | OL | radical (adj) | 31.7 | AWL | ethnic group | 52.3 | AWL |
| prevention | 34.8 | K1 | extended family | 30.7 | | first-born (n) | 51.1 | |
| hierarchy | 27.7 | AWL | prediction | 27.7 | AWL | anti- (prefix) | 44.9 | OL |
| alternate (vb) | 27.4 | AWL | conventional | 25.7 | AWL | format (n) | 42.0 | AWL |
| optimistic | 22.1 | OL | oak tree | 21.8 | | hardship | 41.9 | K1 |
| problematic | 22.1 | K1 | home improvement products | 20.8 | | eye scan | 36.0 | |
| gourmet (adj) | 14.2 | OL | reunion | 20.8 | OL | density | 23.3 | OL |
| runster | 8.0 | NW | hormone | 19.8 | OL | generosity | 14.9 | K2 |
| ecution | 8.0 | NW | life expectancy | 11.9 | | brag (vb) | 14.0 | OL |
| mengel | 7.1 | NW | ecutian | 5.0 | NW | runster | 10.2 | NW |
| Freud | 6.2 | OL | runster | 5.0 | NW | ecutian | 4.6 | NW |
| wuggy | 6.2 | NW | mengel | 3.0 | NW | pregime | 3.5 | NW |
| pregime | 6.2 | NW | pregime | 2.0 | NW | wuggy | 2.2 | NW |
| census | 5.3 | OL | wuggy | 0.0 | NW | mengel | 0.0 | NW |

K% = % of learners reporting the item known (item facility in percent)

*f* = item frequency for that lemma (from Web Vocabprofile, Cobb, 2006)

K1 = first thousand words

K2 = second thousand words

AWL = Academic Word List

OL = off list, not in K1, K2, or AWL

NW = nonword

Blank = frequency not obtained