Exploring the lexical challenge of the TOEIC® *Bridge*

Raymond Stubbe
Kyushu Sangyo University

Reference data:
Stubbe, R. (2010). Exploring the lexical challenge of the TOEIC® *Bridge*. In A. M. Stoke (Ed.), *JALT2009 Conference Proceedings*. Tokyo: JALT.

A yes/no vocabulary checklist test of 120 items was developed by randomly selecting words from the Reading Sections of two official *TOEIC® Bridge Practice Tests* (Ashmore et al., 2007). First and second year university students (N = 234) completed the yes/no test during regularly scheduled classes. Multiple regression analysis was performed on the test results, on the following five factors: *loanword* status (is the item a cognate in Japanese?); number of *letters*; number of *syllables*; *GSL frequency* (West, 1953); and Vocabprofile (Cobb, n.d.) *word level* status (i.e., 1st 1K, 2nd 1K, AWL (academic word list), or off-list). It was found that the first 2 factors, *loanword* status and number of *letters* had the highest correlations with test scores. These results suggest further research into *loanwords* as a lexical focus when preparing students for the *TOEIC® Bridge* may be warranted.

TOEIC® Bridge公式問題集(Ashmore et al., 2007)に掲載されている、2つの模擬テストのリーディングセクションから無造作に選ばれた120語から成るyes/no語彙知識確認テストを作成し、大学1年生と2年生(234名)を対象に、授業中にテストを実施した。テスト結果を単語が外来語であるか否か、単語の文字数、単語の音節の数、GSL(West, 1953)における単語の使用頻度、そしてVocabprofile (Cobb, n.d.)における単語レベル(最初の1,000語、2番目の1,000語、academic word list、またはoff-list)の5つの変数について、重回帰分析行った。その結果、外来語であるか否かと、単語の文字数の2つの変数とテストの結果に、一番高い相関関係が認められた。これらの結果から、英語からの外来語に注目した語彙学習プログラムの、TOEIC® Bridge対策への効果をさらに研究する価値があると考えられる。

Helping students to prepare for upcoming examinations is one of the primary responsibilities of many language teachers. One such test, of increasing importance, is the TOEIC® *Bridge* test. By way of example, at the university where this writer is affiliated, first and second year students take two TOEIC® *Bridge* tests annually, one just before the school year commences and one following its conclusion in mid-January. The initial test scores are used for placement purposes within the English program, while the latter test results account for 30% of the students' second semester final grade. Consequently, English teachers at this institution are compelled to assist students to improve their TOEIC® *Bridge* abilities by focusing on either listening or reading skills. Another component of this English program which has been receiving increasing attention is lexical development. Selecting which vocabulary should be taught in these English classes to best improve the students' TOEIC® *Bridge* scores is the focus of this study.

Test design

In line with a student-centered teaching philosophy, it was decided to consult the students themselves concerning which TOEIC® *Bridge* vocabulary items they may need assistance with. To this end, a 120 item yes/no vocabulary checklist (paper) test was developed. It was based on a random sampling of the lexical items contained in the Reading Sections (Parts 4 and 5) of two official TOEIC® *Bridge* Practice Tests (Ashmore et al., 2007), as copies of real TOEIC® *Bridge* tests are not available. First and second year students (n=234) completed the yes/no test during regularly scheduled classes.

Compared to other vocabulary test designs, yes/no tests allow for the testing of the greatest number of words, while placing the least demands upon test-takers and administrators (Anderson & Freebody, 1983). In this format, test-takers are presented with a series of words in isolation and simply indicate whether they know each item or not (Read, 1993, 2007). A *yes* response is known as a "hit", and a *no* response is a "miss". Unlike most yes/no test research (e.g., Anderson & Freebody, 1983; Huibregtse, Admiraal, & Meara, 2002; Meara & Buxton, 1987), pseudowords were not included in this 120 item test for a number of reasons. Firstly, as this test is being used only for research, as opposed to grading or placement purposes, it was felt that pseudowords, which only provide a measure of over-estimation of vocabulary size/ability, would contribute little to the present study. Secondly, to avoid over-taxing the students, it was decided to maximize the number of real words by not including any pseudowords. Finally, a number of other researchers (Harrington & Carey, 2009; Mochida & Harrington, 2006; Shillaw, 1996) have questioned the necessity of pseudowords in this test format, as there appears to be "little difference between performance on yes/no tests containing words and pseudowords, and those containing (real) words alone" (Mochida & Harrington, 2006, p. 92). Additionally, as "underestimation (not overestimation) more strongly influenced (Y/N) test scores" for a similar group of students (Stubbe, Stewart & Pritchard, 2010, p. 19), the inclusion of pseudowords was deemed unnecessary. The aim of the present study is to determine what insights (if any) can be gained from a yes/no test of the students' knowledge of a sampling of the words contained in two TOEIC® *Bridge* Practice Tests, as well as any implications these insights may have for lexical instruction aimed at helping students prepare for the TOEIC® *Bridge*.

Methods

Participants

The participants in this study were first and second year students (n = 126, 108 respectively) at a private university in southern Japan. Participants had previously been placed into one of three class levels according to their placement TOEIC® *Bridge* scores (low: 70-99; medium: 100-119; high: 120 +). Two first-year and three second-year classes per class level completed the yes/no test during regular class-time between October 5 and 16, 2009.

Yes/No test preparation

The individual words in two sections of the TOEIC® *Bridge* 公式ガイド & 問題集 (Official Guide and Workbook) (Ashmore et al., 2007) were selected for inclusion in this yes/no test. These sections were the Reading Practice Section (pp. 52-63), and the Reading Section – Parts 4 and 5 of Practice Test 4 (pp. 78-89), and contained 959 words. It was decided not to include the listening sections as vocabulary knowledge has a much higher correlation with reading over listening tasks (Anderson & Freebody, 1981; Ellis & Beaton,1993). All test instructions, headings, proper nouns, punctuation, numbers, and symbols were removed, as were any words common to this university's 500 Word Vocabulary

Program. The remaining 381 words were then pasted into Vocabprofile (Cobb, n.d.) to sort them into four groups: 1K, 2K, AWL, and off-list. The first group, 1K, is comprised of the most frequent 1,000 English words, as determined by a frequency list derived from the British National Corpus (BNC) by Paul Nation (Daulton, 2003). The second group, 2K, is comprised of the second most frequent 1,000 English words, from that same frequency list. *AWL* stands for *Academic Word List*, a listing of 570 word families from academic texts (Coxhead, 2000). *Off-list* refers to words not contained in the first three lists. These words were then assigned random numbers and the first 30 in each of the four word groups were selected for inclusion. In one instance, multiple versions of a word were randomly selected, i.e., *identification, identifies*, and *identity*. In this case, only the base word, *identify*, was included in the test, and the remaining two words were replaced by the next two randomly selected words. Generally, the form of the words found in the Practice Test was kept in the yes/no test, e.g., *published, strategies, decorating*, as these are what test-takers would actually encounter. In the resulting 120-item yes/no test, words were ordered alphabetically, and then placed into three columns of 40 items each. To avoid the possible influence of sequence effects, a second version of this yes/no test was created by reversing the word order in each of the three columns. Hence a two version yes/no 120-item paper test was created. Each version (v.1 and v.2) contained the following instructions: "For the words you know the meaning of, fill in the circle to the right." ("鉛筆またはシャープペンシルを使用しなさい。あなたが意味を知っている単語について、右側のマークを塗りつぶしなさい。")

Procedure

The participating six teachers distributed the tests to their students during regular class-time, in such a way that every student had an equal chance of getting one of the two versions. Generally, students completed the test in 8 to 15 minutes. Forms were then collected, separated by version, scanned, and converted into Excel files for analysis. In addition to reporting means and standard deviations, a Rasch analysis provided item and person separation and reliabilities. A paired t-test of the Rasch Measures for v.1 and v.2 was conducted to check for equivalency between the two test versions. Pooled t-tests were conducted to check for significant differences between the various word groups (1K, 2K, AWL and off-list). To account for the variance in the yes/no test results, a number of regression analyses were undertaken on the following factors: *loanword* status, number of *letters* and *syllables*, *GSL frequency level*, and *Vocabprofile word level*. Finally, paired t-tests compared the results of the top 50% of test-takers to the bottom 50%.

Results and discussion

Version 1 was completed by 120 students and v.2 by 114. The greatest number of words checked as known by one student was 113 of 120, the lowest was 18. The mean number of hits per student was 69.52 words (57.93%), with a standard deviation (*SD*) of 19.79 words. Results for v.1 and v.2 are displayed in Table 1. Both versions of this yes/no test were found to be highly reliable, with Cronbach's alpha = .96.

Table 1. Mean hits, SDs, and test reliability for v.1, v.2, and combined (n = 234)

Test version	n	Mean	*SD*	Cronbach's alpha
1	120	68.88	19.84	.96
2	114	70.46	20.27	.96
Combined	234	69.52	19.79	.96

Two participants' results in v.1 were found to be outliers, likely because they did not take the test seriously enough. In both cases, their results were far below the mean for their level. Table 2 reports the same statistics as Table 1, with these two outliers removed. Although these two outliers had little effect on the analysis thus far, they would seriously skew any finer classical analysis such as t-tests and regression analyses (discussed below), so their removal from the data set was deemed necessary.

Table 2. Mean hits, SDs, and test reliability for v.1, v.2, and combined, outliers removed (n = 232)

Test version	*n*	Mean	*SD*	Cronbach's alpha
1	118	68.58	19.30	.96
2	114	70.46	20.27	.96
Combined	232	70.00	19.74	.96

A Rasch analysis was also performed on v.1 and v.2, as well as on the full combined results. As the means and *SD*s are almost identical to Table 1, Table 3 only reports the reliability and separation for persons (participants) and items (words). Again, reliabilities are very high (.96 or higher) and the participants were separated into five distinct levels. The separation of the 120 words into six distinct levels per version is even stronger. The combined analysis was repeated after removing the two outliers.

Table 3. Rasch analysis results: separation and reliability

Test version	*n*	Separation	Reliability
1	120 Persons	5.01	.96
	120 Items	6.96	.98
2	114 Persons	5.26	.97
	120 Items	6.42	.98
Combined	234/232* Persons	5.11/*5.05**	.96/*.96**
	120 Items	9.77/*9.61**	.99/*.99**

Note: Italics denote figures for n = 232, two outliers deleted.

Rasch analysis was also used to check that the two yes/no test versions were equated. After removing the misfitting persons, a Rasch analysis was reapplied to uncover any misfitting items. The words *tech* and *gulf* were strongly misfitted, and so were removed from the data sets. The removed persons were then replaced, and a third analysis of v.1 and v.2 was undertaken. By comparing the item difficulty of v.1 and v.2, it was found that there were no significant differences between the two versions. A paired t-test of the Rasch measures for v.1 and v.2 confirmed that no significant difference existed between the two means (t = .87, df = 117, p = .386).

As mentioned above, this test was taken by students in three ability levels. As expected, results improved with ability level. Table 4 shows the means and *SD*s for the low, medium (med), and high levels. A simple regression analysis produces an R^2 value of 43.2% (t = 13.22, df = 230, $p \leq .0001$), meaning that 43.2% of the variance in test scores can be explained by variance in student level.

Table 4. Yes/No test results by student level for 120 words

Level	*n*	Mean (word hits)	*SD*
Low	88	54.43	16.43
Med	73	73.65	13.90
High	71	85.58	13.67

Table 5, which breaks down the results by year, reveals a rather unexpected finding. The low second-year students outperformed their first-year counterparts, while the reverse situation occurred in the two higher levels. An explanation for this situation is beyond this writer, as generally second-year students tend to be more apathetic towards such activities in this institution and low level students tend to be the most apathetic.

Table 5. Yes/No test results by student level by year for 120 words

Level	*n*	Mean (word hits)	*SD*
low 1st year	40	47.98	15.14
low 2nd year	48	59.81	15.63
med 1st year	43	78.16	13.67
med 2nd year	30	67.17	11.64
high 1st year	42	88.93	11.19
high 2nd year	29	80.72	15.57

An analysis of the 4 word groups (1K, 2K, AWL, off-list, n = 30 each) is shown in Table 6. As might be expected, the students knew the 1K words much better than the AWL and off-list words, but not significantly better than the 2K words.

Table 6. Results by word group for words and participants

Word group	Word mean (n = 30 each)	*SD*	Participant mean (n = 232)	*SD*
1K	21.82	5.11	169.37	62.47
2K	21.03	5.11	163.37	63.98
AWL	13.27	5.81	103.33	70.39
Off-list	13.53	5.24	105.33	68.34

Pooled t-tests comparing the different word group pairs are presented in Table 7. No significant differences were found between the 1K and 2K pair, or the AWL and off-list pair. All other pairs were significantly different.

Table 7. Pooled t-tests of the four word groups (*df* = 115)

Word group	*t* value	*p* value
1K-2K	1.59	.097
1K-AWL	17.1	<.001
1K-Off-list	17.47	<.001
2K-AWL	15.46	<.001
2K-Off-list	15.82	<.001
AWL-Off-list	0.5089	.611

In order to start accounting for the variance in the yes/no test results, a number of regression analyses were undertaken. As mentioned, a simple regression analysis of hits on ability level produced an R^2 value of 43.2%. Breaking hits up into the 4 word groups, a multiple regression analysis for the four word groups on the three student levels was performed. Although R^2 increased to 43.8%, only results for the 1K ($t = 2.76$, $df = 26$, $p = .006$) and the AWL group ($t = 2.43$, $df = 26$, $p = .016$) were significant. Results for the 2K words ($t = .58$, $df = 26$, $p = .561$) and off-list words ($t = .61$, $df = 26$, $p = .545$) were not significant. So only 43.8% of the variance in test score by word group could be accounted for by variance in word group.

Loanwords and letters

A look at the results for individual words was more revealing. As can be seen in the Appendix (a listing of the 120 words and their scores) *piano*, an off-list word, received the maximum number of hits ($n = 232$). This led to an investigation into loanwords, as *piano* is a very common English loanword in Japanese. Daulton (2003) compiled a list of loanwords containing 1,777 word-types, found within the BNC 3000 word families, plus 35 additional academic words. According to Daulton, "About 45.5 percent of the 3000 most-frequent word families in English have correspondences to common Japanese loanwords" (Daulton, 2003). Perhaps surprisingly, it was found that 51.7% of the words on this yes/no test are on Daulton's (2003) English Loanwords in Japanese list, while 45.8% of the 871 TOEIC® *Bridge* Practice Test *non-proper noun* words are on that same list. Adding the off-list words *soccer* and *hiking* to that loanword list increases the above figures to 53.3% and 42.8%, respectively. Sixty-four of the 120 items on the yes/no test are loanwords, and the remaining 56 are non-loanwords.

A multiple regression analysis was run on the following five factors: loanword (yes or no), number of letters, number of syllables, word group (1K, 2K, AWL, or off-list), and frequency according to the GSL (West, 1953; "a set of 2,000 words selected to be of the greatest 'general service' to learners of English", listed in order of their frequency; Bauman, n.d.). The number of letters was counted for each word and recorded, as was the number of syllables. The latter was chosen because Perkins and Linnville (1987) reported that the number of syllables is a good predictor of word difficulty. GSL frequency proved slightly challenging as the list (Bauman & Culligan, 1995) contains only the 2,284 most frequent words. Sixty of the 120 items, mostly from the AWL and off-list groups are not included in this GSL list. The Vocabprofile BNC-20 (Cobb, n.d.) was consulted. Twenty-one of the 60 words had frequencies in the 3000 level or higher. As the BNC-20 lists frequencies as 1000, 2000, 3000, etc., the mid-range figures were used (2500, 3500, 4500, etc.) However, 19 words were in the BNC-20 2K level, and 20 words were in the 1K level. As these words do not appear in the GSL list which terminates at 2284, they were all arbitrarily assigned values of 2500. Table 8 shows the means and *SD*s for four of these factors.

Table 8. Means and SDs for loanword, letters, syllables, and GSL frequency

Factor	Mean	*SD*
Loanword	.5333	.501
Number of letters	6.083	2.233
Number of syllables	1.992	.893
GSL frequency	1896.04	1282.13

The multiple regression analysis run on all five factors (loanword, letters, syllables, GSL frequency, and word group)

produced the following regression equation: "Hits = 208.12 + 62.71 loanword – 17.42 letters + 23.30 syllables + 0.01 GSL – 12.80 word group" with an R^2 value of 49.9%, and an F-ratio of 22.7. *Loanwords* had the highest correlation ($t = 6.39$, $df = 114$, $p \leq .0001$); *letters* was second ($t = -4.28$, $df = 114$, $p \leq .0001$); and *syllables* was third, but still significant ($t = 2.31$, $df = 114$, $p \leq .0227$). However, *syllables* was positive in the regression equation, illogically suggesting that number of hits increases with number of syllables (Table 9). Neither *GSL* nor *Word Group p* values were significant.

A second analysis without *GSL* still resulted in a positive correlation for *syllables*. A third analysis without *syllables* resulted in small *t* values for *GSL* and *word group*. A fourth analysis was also carried out without *word group*. In all four multiple regressions, *loanword* consistently had the highest correlations, followed by *letters* in all but the second analysis. Simple regressions were also performed on each of these five factors and *loanwords*, again, had the highest correlations (Table 10).

Table 9. Multiple regression analysis: 5 factors

Number of factors	R^2 value	*f*-ratio	Factor	*t*-ratio	*p*-ratio
5, all ($df = 114$)	49.9%	22.7	loanword	6.39	**≤ .0001**
			letters	-4.28	.0004
			syllables	2.31	**.0227**
			GSL	-1.3	.1951
			word group	-1.85	.0666
4, GSL removed ($df = 115$)	49.2%	27.8	loanword	6.33	**≤ .0001**
			letters	-4.21	**≤ .0001**
			syllables	2.23	**.0280**
			word group	-4.28	**≤ .0001**
4, syllables removed ($df = 115$)	47.6%	26.1	loanword	6.35	**≤ .0001**
			letters	-4.01	**≤ .0001**
			GSL	-1.14	.2574
			word group	-1.78	.0780
4, word group removed ($df = 115$)	41.3%	27.2	loanword	6.32	**≤ .0001**
			letters	-4.45	**≤ .0001**
			syllables	2.25	**.0261**
			GSL	-4.05	**≤ .0001**

Note: *GSL* means GSL frequency, *letters* and *syllables* mean number of letters and syllables, *word group* means Vocabprofile level (1K, 2K, AWL, off-list), numbered according to hits (e.g., 1.01 equals the 1K word with the most hits, 4.30 equals the off-list word with the least hits.

Significant *p*-ratios are in **bold** (alpha = .05)

Table 10. Simple regression analysis: 5 factors ($df = 118$)

Factor	R^2 value	*f*-ratio	*t*-ratio	*p*-ratio
Loanword	24.6%	38.5	6.21	≤ .0001
Letters	22.5%	34.3	-5.86	≤ .0001
Word group	15.2%	21.2	-4.6	≤ .0001
GSL	11.7%	15.6	-3.95	≤ .0001
Syllables	10.7%	14.1	-3.76	.0003

A deeper analysis of the 64 loanwords versus the 56 non-loanwords is somewhat informative. The respective means for hits/word are 168.89 and 97.02, with *SD*s of 57.07 and 69.75

hits. Although loanwords made up 53.33% of the items on the test, they accounted for 10,809 (65.55%) of the hits, whereas non-loan words received 5,433 of the total 16,242 hits (33.45%). A two-sample t-test of students' scores on loanwords versus non-loanwords reveals that the difference between the means is 134.81 ($t = 20.34$, $df = 119$, $p \leq .0001$), confirming that the difference between loanword and non-loanword results is statistically significant.

A comparison of the top 50% student scores against the bottom 50% was also performed. The means for the high group versus the low group were 85.54 and 54.47, with *SD*s of 10.77 and 13.41, respectively (Table 11). A paired t-test comparing these two groups on the full 120 words gives a difference between means of 30.03 ($t = 17.25$, $df = 119$, $p \leq .0001$) confirming that the differences are statistically significant. Comparing their knowledge of loanwords versus non-loanwords, the high level students knew 76.3% (6,332 / 3,591) more loanwords than non-loanwords, while the low group knew 143.1% (4,477 / 1,842) more loanwords than non-loanwords. It appears that loanwords make up a much greater proportion of the vocabulary of these low level learners (Table 12).

Table 11. Mean hits, *SD*s for top 50% of students against bottom 50%

Test version	*N*	Mean	*SD*
Top 50%	116	85.54	10.77
Bottom 50%	116	54.47	13.41

Table 12. Loanword vs. non-loanword hits by top 50% of students against bottom 50%

Students	Loanword hits	% of max. possible loanwords	Non-loanword hits	% of max. possible non-loan-words
Top 50% ($n = 116$)	6332	85.3%	3591	55.3%
Bottom 50% ($n = 116$)	4477	60.3%	1842	28.0%
Total	10809	72.8%*	5433	41.8%*

Note: maximum possible loanwords/group (116 x 64) = 7,424 hits, maximum possible non-loanwords/group (116 x 56) = 6,496 hits.

* Average percentage of the high and low groups

Finally, a similar high-low analysis was performed on number of letters. The means for the high student half versus the low half were 86.69 and 52.66, with *SD*s of 34.90 and 39.98, respectively. Compared to the high group, the lower students' score deteriorated starting at 6 letters/word, and seriously deteriorated at 8 letters (Table 13).

A paired t-test comparing the high group against the low also yields a difference between means of 30.03 ($t = 17.25$, $df = 119$, $p \leq .0001$), again significantly different. Simple regressions on the two groups gives respective R^2 values of 17.8% and 24.4%, with F-ratios of 25.5 and 38 ($t = -5.05$ and -6.17, $df = 118$, $p \leq .0001$). The low group's results correlate moderately better with number of

Table 13. Number of letters per word by top 50% of students against bottom 50%

Letters/Word	2	3	4	5	6	7	8	9	10	11	12
n (words)	2	8	26	15	22	19	8	10	5	3	2
possible hits	464	1856	6032	3480	5104	4408	1856	2320	1160	696	464
high (hits)	231	715	2563	1541	1651	1463	465	613	352	207	28
low (hits)	224	603	1886	1172	923	779	169	297	159	66	5
% higher	3.13	18.6	35.9	31.5	78.9	87.8	175.2	106.4	121.4	213.6	460.0

Note: *n* represents number of words with *x* letters, *possible hits* represents maximum possible hits (*n* - words x 232), *% higher* equals high score / low score – 1 x 100.

letters (-.421 versus -.494, high/low respectively). It appears as if number of letters is also a stronger factor for the lower students.

The implications of these results for lexical instruction to assist students improve their TOEIC® *Bridge* scores are unclear. Perhaps they merely confirm the suspicions of many second language teachers: word difficulty increases with number of letters, and that any words commonly appearing as cognates in the L1 are more likely to be recognized. However, that low-level students' word recognition deteriorates after five letters (and especially after seven) is worth bearing in mind. Also, although these students do know a much higher proportion of loanwords than non-loanwords (72.8% versus 41.8%), focusing lexical instruction on the former may still benefit students. Given that 43% of the lexical items on these practice tests are loanwords, and that a good number of loanwords in Japanese have acquired meanings and uses different from those found in native English, (e.g., *smart*, *unique*, and *mansion*, Ishii, 2005) teaching students these differences should help to avoid confusion during testing situations. Also as the low students knew only 60.3% of loanwords whereas the high students knew 85.3%, it is possible that the low-level students have more difficulty in recognizing loanwords which they already know in L1. Expanding the low students' ability to recognize both written and spoken forms of such loanwords may also help increase TOEIC® *Bridge* scores. Further research is required to confirm these speculations.

Conclusion

This study has been an enquiry into the vocabulary knowledge of one university's first and second year students regarding a sampling of words taken from the Results Section of two TOEIC® *Bridge* Practice Tests. Although the convenience sampling used in this research makes it difficult to generalize from the findings, and the four words *gulf, tech, mark* and *projects* should be replaced, these findings may be of interest to vocabulary researchers working with Japanese university students. The developed yes/no test had good person and item reliability and separation, with students knowing 58% of the 120 words. The 1K words were known much better than the AWL and

off-list words, but not significantly better than the 2K words. Multi-regression analysis suggested that two factors – whether a word is a loanword in Japanese, and the number of letters – seem to strongly influence word recognition, especially for the lower-level students. These lower students knew 143% more loanwords than non-loanwords, while the higher students knew 76% more loanwords. These results suggest that further research focusing on loanwords is warranted to help low-level students better prepare for the TOEIC® *Bridge*.

Acknowledgments

I would like to thank Trevor Holster for all of his assistance with the Rasch analyses, and Miki Tokunaga for translating the abstract into Japanese.

Bio data

Raymond Stubbe holds an M.A. in Applied Linguistics and TESOL from the University of Leicester, England; and a B.Ed. from the University of Victoria, Canada. He is a lecturer at Kyushu Sangyo University. His research interests include vocabulary acquisition and testing. <raymondstubbe@gmail.com>

References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-177). Newark, DE: International Reading Association.

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

Ashmore, E., Carter, E., Duke, T., Hauck, M., Locke, M., & Shearin, R. (2007). *TOEIC® Bridge*公式ガイド&問題集. Princeton, NJ: Educational Testing Service.

Bauman, J. (n.d.). *About the General Service List*. Retrieved Nov. 5, 2009, from http://jbauman.com/aboutgsl.html.

Bauman, J., & Culligan, B. (1995). *The GSL*. Retrieved June 11, 2009, from http://jbauman.com/gsl.html.

Cobb, T. (n.d.). *Web Vocabprofile*. Retrieved June 12, 2009, from http://www.lextutor.ca/vp/ Coxhead, A. (2000). A new "Academic Word List". *TESOL Quarterly, 34*, 213-238.

Daulton, F. E. (2003). List of high-frequency baseword vocabulary for Japanese EFL students #2. *The Internet TESL Journal, 14*. Retrieved from http://iteslj.org/lists/Daulton-BasewordVocab2.html.

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning, 43*, 559–617.

Harrington, M., & Carrey, M. (2009). The on-line yes/no test as a placement tool. *System, 37*, 614-626.

Heatley, A., & Nation, P. (1994). Range [Computer software]. Available from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx.

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes/no vocabulary test: Correction for guessing and response style. *Language Testing, 19*, 227-245.

Ishii, T. (2005). Diagnostic tests of vocabulary knowledge for Japanese learners of English. Unpublished doctoral dissertation. University of Nottingham, UK.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*, 142-151.

Mochida, A., & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary. *Language Learning, 23*, 73-98.

Perkins, P., & Linnville, S. E. (1987). A construct definition study of a standardized ESL vocabulary test. *Language Testing, 4*, 125-141.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Learning, 10*, 335-371.

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 modeling to yes/no vocabulary tests. *Vocabulary Acquisition Research Group*, University of Wales Swansea. Retrieved June 10, 2009, from: http://www.lognostics.co.uk/vlibrary/index.htm.

Stubbe, R., Stewart, J., & Pritchard, T. (2010). Examining the effectiveness of pseudowords in yes/no vocabulary tests for low level learners. *Kyushu Sangyo University Language Education and Research Center Journal, 5*, 5-23.

West, M. (1953). *A General Service List of English Words*. London: Longmans Green.

Appendix

120 words sorted by word group and score

1K	Score	2k	Score	AWL	Score	Off-list	Score
game	229	coffee	229	computers	215	piano	232
he	228	bus	227	goal	215	soccer	217
we	227	club	226	job	215	airport	205
old	226	yellow	226	team	213	television	198
music	226	ice	222	project	205	hobby	195
last	225	mail	220	network	200	museum	172
seven	220	orange	219	area	183	math	169
what	219	lunch	218	feature	137	hiking	161
they	218	key	214	projects	133	tiger	159
left	218	slow	210	plus	130	apartments	155
total	209	telephone	209	site	126	protector	144
soft	205	bicycle	208	recover	123	jet	126
some	204	advice	202	manuals	122	heater	123
popular	204	sport	202	credit	121	traffic	110
heavy	201	ocean	197	corporation	120	gulf	103
ride	183	fashion	183	creates	92	decorating	97
beauty	178	luck	169	published	76	resort	84
mark	175	excite	165	issues	62	client	68

1K	Score	2k	Score	AWL	Score	Off-list	Score
marry	168	medicine	165	locate	58	laundry	61
service	159	engine	140	attach	57	helpfully	53
round	155	lend	134	instruction	54	updates	49
interest	146	wool	130	responses	51	patients	48
actor	144	tent	109	identify	45	zip	46
such	138	flat	106	annual	37	viewers	44
introduce	129	temperature	99	strategies	32	tech	43
purpose	77	repair	93	construction	27	fares	26
profit	46	earn	79	purchase	22	basement	25
nor	42	postpone	54	residents	15	resume	19
concern	41	instrument	29	predicted	8	slippery	18
length	41	barely	17	**intermediate**	6	subscribe	10

Note: All loanwords are underlined; highest and lowest scores are in **bold**.