Who check more pseudowords, low-level or high-level students?

Raymond Stubbe
Kyushu Sangyo University

Chris O'Sullivan
Kitakyushu University

Jeremy Boston
Tokyo University of Foreign Studies

Mathew Porter
Hiroshima Shudo University

Richard Grumbine
Ariake National College of Technology

David Latz
Shimonoseki City University

Pseudowords, or non-real words, have been used in Yes/No vocabulary tests for nearly three decades (Anderson & Freebody, 1983). Recent research has suggested that their inclusion can cause problems for low-level English learners (Meara, 1996; Cameron, 2002). The aim of this study is to determine whether low-level Japanese university EFL learners have more pseudoword checks, or false alarms, than their high-level counterparts. Thirty-four classes in six different universities participated (n = 738). Students took two separate Yes/No vocabulary tests of 48 real words each. The second test also included 32

pseudowords. Whereas the lower group of students had a false alarm rate of 3.66%, the higher group had a false alarm rate of 4.68%. These results suggest that, contrary to expectations, the pseudowords were not significantly more difficult for the lower-level students.

　ここ30年間、疑似語、つまり実在しない語彙は、Yes/No語彙力テストに使用されてきた（Anderson & Freebody, 1983）。最近の研究では、このようなテストに疑似語を加えると、習熟度の低い英語学習者に困難を招く可能性があると指摘されている（Meara, 1996; Cameron, 2002）。本研究の目的は、習熟度の低い日本人大学生EFL学習者が習熟度の高い者よりも、疑似語に対してYesと答える、つまりfalse alarmとなる場合が多いか否かを検証することである。本研究は、6大学、34クラスの学生（738名）を対象に実施し、学生は、実在する語彙48語をそれぞれ含んだ2種類のYes/No語彙力テストを受験した。2種類目のテストには、実在する語彙48語に加え、疑似語32語も含んだ。習熟度の低い学生グループのfalse　alarm率が3.66％であったのに対し、習熟度の高い学生グループのfalse　alarm率は4.68％であった。この結果から、予想に反して、疑似語は習熟度の低い学生にとって著しく難解ではなかったと考えられる。

THE Yes/No (Y/N) vocabulary test has been and remains a popular format for measuring students' receptive lexical knowledge. In these tests, students are presented with a list of individual words and instructed to indicate the ones they know the meanings of. Yes/No tests are simple to construct and administer; plus they allow for the testing of over twice as many words as comparable multiple-choice tests (Anderson & Freebody, 1981). These tests can be used to measure students' overall vocabulary size or to determine specific vocabulary knowledge in preparation for upcoming language units or activities. "Despite its simplicity, the Yes/No format has proved to be an informative and cost-effective means of assessing the state of learners' vocabulary knowledge, particularly for placement and diagnostic purposes" (Read, 2007, 112-113). Although these tests may be useful for determining a person's breadth of vocabulary knowledge, they are criticized for their inability to measure the depth of lexical knowledge (Chapelle, 1998; Ishii & Schmitt, 2009; Laufer & Goldstein, 2004; Read, 2000).

As these tests rely solely on "self reporting", the actual knowledge of the students cannot be verified. Unfortunately, students' self-reporting of whether they know a word or not is sometimes a poor indicator of their actual vocabulary knowledge. Chall and Dale (1950, cited in Anderson & Freebody, 1981) reported an average tendency to overestimate word knowledge at about 11%. To compensate for the potential of students to claim knowledge of words they do not actually know the meaning of (overestimation), pseudowords (or non-words) were introduced to the vocabulary test (Anderson & Freebody, 1983). This use of pseudowords in Y/N tests has remained popular through present-day versions. In these tests, knowledge of a real word is known as a "hit", while claiming knowledge of a pseudoword is a "false alarm". Not claiming knowledge of a real word is labeled a "miss" and not claiming knowledge of a pseudoword is a "correct rejection". Claiming knowledge of words that do not exist is seen as evidence of falsely claiming knowledge of real words (overestimation).

Although pseudowords may counteract the problem of overestimation, their use in Y/N tests is not without difficulties. Creating adequate pseudowords is itself a challenge as they "should respect the phonotactic and morphological rules of the target language" (Eyckmans, 2004, p. 27), but "not bear meaning" (Huibregtse, Admiraal, & Meara, 2002, p. 227). There are two methods of creating pseudowords. One method, pseudo-derivatives, entails adding a prefix or suffix to a real word, so *loyal* becomes *loyalment*. In the second method, one or two vowels and/or consonants are substituted in a real word, so *boy* becomes *poy* (Anderson & Freebody, 1983). A second potential difficulty with pseudowords is that they may influence some students to become too conservative with their vocabulary knowledge by not checking words that they actually do know (underestimation) (Eyckmans, 2004; Huibregtse, et al., 2002; Mochida & Harrington, 2006).

A further weakness that has been found with the Y/N test is that it "does not perform well with low-level learners, who respond unpredictably to the pseudowords" (Beeckmans, Eyckmans, Janssens, Dufranne, & Van de Velde, 2001, p. 240). Apparently, low-level students "often misread items in the test in unpredictable ways", which results in high false alarm rates (Meara, 1996, p. 42). Cameron similarly found that pseudo-words created problems for her lower, English as an Additional Language students, "many of whom had scores heavily reduced by the number of non-words they checked as 'known'" (Cameron, 2002, p. 159). False alarm rates, or the percentage of actual false alarms divided by the maximum number possible, have been reported from 5% through 25% (Cameron, 2002; Eyckmans, 2004; Mochida & Harrington, 2006). However, in the pilot to this study, we found that for the low-level participants (with TOEIC® scores around 230) the false alarm rate was a low 4% (Stubbe, Stewart & Pritchard, 2010).

Aims

The specific aim of this research was to ascertain whether or not low-level Japanese university students check more pseudowords in Y/N vocabulary tests than their high-level counterparts. It was hypothesized that students from lower-level English ability universities would have significantly more false alarms compared with their counterparts from higher-level institutions.

Sample

Thirty-four English classes from six different universities in a variety of locations in Japan participated in the study. English ability measures ranged from 230 to 800 on the *TOEIC® test*. Two groups were created in each of the participating universities, one for each of the two versions of the two Y/N vocabulary tests, discussed below.

Method

This study replicates the one described in Mochida and Harrington (2006), with the following exception. The test instructions for the pseudoword tests did not advise the students that they may be tested on some of the real words in the future. It was felt that such a warning might cause the students to become too conservative and underestimate their word knowledge. The test instructions for the pseudoword tests were:

- Use a pencil.
- For the words you know the meaning of, fill in the circle to the right.
- Some of the words do not exist in English.
- 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。
- いくつかの単語は、英語で存在しないものもあります。

Additionally, whereas Mochida and Harrington (2006) tested advanced students, this study focused on students with a wide range of ability levels.

An initial wordlist was created for the pilot study by randomly selecting 96 words from a variety of sources including: the Vocabulary Levels Test (Nation, 1990), the Vocabulary Size Test (Nation & Belgar, 2007), and Webb's translation test (Webb, 2008) in the following word frequency ranges: 0-1000, 1001-2000, 2001-3000 (50%, 33.33%, 16.67%, respectively). Word frequencies were based on the General Service List (GSL) (West, 1953), which is available on-line at http://jbauman.com/aboutgsl.html (Bauman & Culligan, 1995). Six of the words from the pilot study were found to be too easy and were replaced by six lower frequency words. This new list contained 96 real words whose frequencies, based on the BNC-20 listing (a computer program

that performs lexical text analysis based on the British National Corpus (Nation & Cobb, nd)), are shown in Table 1. The 10K word is "amongst" which is listed in the GSL as number 2226 and the 7K word is "weekday" which is listed as number 2255 (West, 1953; Bauman & Culligan, 1995).

Table 1. Word frequency list

Frequency level	Number of words (total 96)
1K, or most frequent 1000 words	48
2K, or second most frequent 1000 words	27
3K	10
4K	2
5K	4
6K	3
7K	1
10K	1

These 96 words were sorted into two equal size lists of roughly equal difficulty, based primarily on the results of the pilot study, to create Y/N tests 48A and 48B. A second series of Y/N tests was created by adding 32 pseudowords, discussed below, to each of the 48 item tests, thereby creating tests 80A, and 80B. In each university two groups of roughly equal English ability were created. The first group received the 48A then the 80B Y/N tests, with the other group receiving the 48B then the 80A tests. Both tests were given on different days. Any students or classes tested on the same set of real word tests twice (either 48A and 80A; or 48B and 80B) were disqualified from participating in the study and all of their results were deleted from the data. Unfortunately, this did occur with 64 of the 816 test-takers. Additionally, another 14 students did not read the test instructions on the second Y/N test, and intentionally checked the pseudowords and other unknown words. As recognizing pseudowords may be a different construct than receptive vocabulary knowledge, these results were also deleted.

The pseudoword-list was prepared using Anderson and Freebody's (1983) *substitution method* for 32 words from the following levels: 1K - 50%; 2K - 33.33%; 3K - 16.67%. A list containing all 96 real words and the 32 pseudowords was pre-tested on six native English speakers, to check that the pseudowords were identifiable as non-words. Two pseudowords caused trouble for a few test-takers, and were subsequently replaced. The adapted list was re-tested by four different native English speakers, and found to be acceptable. Results of the pilot study suggested that four of the pseudowords - *fomul, oligan, steck* and *thrait* were problematic and were subsequently replaced by *fimul, iligan, sweck* and *thait.* Unfortunately, three other pseudowords were also inadvertently changed, possibly by the spell checker - *thraugh, troet,* and *tuught* became *thraegh, troit,* and *tought.*

Results

Y/N Tests

To begin the analysis, the pseudoword results were removed from the data, allowing for direct comparisons of the real word results. The pseudoword results will be presented in the next section. Summary results of the four forms (48A and B, 80A and B) can be seen in Table 2. Students knew an average of 33.56 of the 48 real words on each test (69.9%), with standard deviations (SD) between 8.45 and 9.02. Real word hits ranged from a low of 4 by two participants in test 48B through 48 hits by seven students in tests 48A, 48B and 80A.

Table 2. Summary statistics of results for all test forms

test	n	mean	SD
48A	358	34.14	8.45
80A	391	32.40	8.88
48A+80A	749	33.23	8.67
48B	403	33.31	8.54
80B	347	34.56	9.02
48B+80B	750	33.88	8.76

Note: Only the results of the 48 real words in 80A and 80B were analyzed here.

Table 3 summarizes the means and SD for each of the six universities, ranked from highest to lowest. Students in the top institution knew approximately 62.6% more words than their counterparts in the lowest university (43.78 / 26.93). Additionally, the lower three universities had a much greater range of variance. This ranking of universities was confirmed by a subsequent multiple-choice test of the same 96 items, which was given in order to calculate the amount of under- and over-estimation.

Table 3. Summary of universities ranked high through low on all 4 Y/N tests (48 real words each)

university	number of tests taken	mean	SD
1	157	43.78	3.55
2	355	38.59	4.32
3	114	36.82	4.51
4	188	35.76	6.14
5	371	28.13	8.21

university	number of tests taken	mean	SD
6	314	26.93	7.68
all universities	1499	35.55	8.74

Pseudowords

Analysis of the two pseudoword tests (80A and 80B) indicated the students had a total of 985 false alarms, a mean of 1.33 false alarms per student (see Table 4). The total false alarm rate was 4.17%. A breakdown of false alarms by university is provided in Table 5. Contrary to expectations, university 1 (from Table 3, which also had the highest TOEIC® scores) had the highest false alarm rate (1.55), and university 5 (the second lowest university in Table 3) had the lowest false alarm rate (1.21).

Table 4. Pseudoword false alarms (f. a.)

test	n	false alarms	f. a. mean	f. a. rate*
80 A	391	504	1.29	4.03%
80 B	347	481	1.39	4.34%
total	738	985	1.33	4.17%

n = number of students
f. a. = false alarm
* false alarm rate = false alarms / max. number of possible false alarms (eg. 504 / 391 x 32 pseudowords)

Table 5. False alarms by university

university	n	false alarms	f. a. mean	f. a. SD	f. a. rate*
5	189	229	1.21	1.67	3.79%
3	56	68	1.21	1.12	3.79%
2	175	227	1.30	1.59	4.05%

university	n	false alarms	f. a. mean	f. a. SD	f. a. rate*
6	155	216	1.39	2.27	4.35%
4	87	127	1.46	1.93	4.56%
1	76	118	1.55	2.58	4.85%
overall	738	985	1.33	1.86	4.17%

n = number of students
f. a. = false alarm
* false alarm rate = false alarms / max. number of possible false alarms

university	n	false alarms	f. a. mean	f. a. SD	f. a. rate*
4	87	127	1.46	1.93	4.56%
overall	736	945	1.29	1.56	4.01

n = number of students
f. a. = false alarm
* false alarm rate = false alarms / max. number of possible false alarms

Further examination revealed two extreme outliers, one each from the highest and lowest institutions. From the highest university, one student reported 21 pseudowords, over 7.54 SDs above the mean of 1.55 false alarms per student for that institution. He also reported the same number of real words (21, or 43.75%) on the same test while receiving 98% on the subsequent multiple-choice test. A second student, this time from the lowest university, reported 19 pseudowords, over 7.73 SDs above that institution's mean of 1.39 false alarms. This student also checked 25 real words on this Y/N test (52.08%) and received 74 % on the multiple-choice test. Removing these 2 outliers from the data (as suggested in De Veaux, Velleman & Bock, 2008, p. 86) decreases the overall false alarm mean from 1.33 to 1.29 and the false alarm rate from 4.17% to 4.01% (see Table 6).

Almost 30% of the 738 students who took these tests reported no false alarms whatsoever. Over 70% of students reported only 0 or 1 false alarms (see Table 7). In a High-Low analysis based on the participant results of the real words portion of the Y/N tests, the high-half of participants was composed primarily of students from the higher universities (69.38%) and had a mean false alarm rate of 4.68%. The low-half was primarily composed of students from the three lower institutions (86.18%) and had a false alarm rate of 3.66%, including the two outliers. It appears as if the original hypothesis may have been disproven – lower-level students do not generally check more pseudowords than their higher-level counterparts.

Table 6. False alarms by university, less two outliers

university	n	false alarms	f. a. mean	f. a. SD	f. a. rate*
5	189	229	1.21	1.67	3.79%
3	56	68	1.21	1.12	3.79%
6	154	197	1.28	1.78	4.00%
1	75	97	1.29	1.25	4.04%
2	175	227	1.30	1.59	4.05%

Table 7. False alarms counts

number of false alarms	count*	number of false alarms	count*
0	219	8	2
1	312	9	1
2	124	11	1
3	35	12	1
4	18	13	1
5	8	14	2
6	5	19	1
7	7	21	1

number of false alarms	count*	number of false alarms	count*
total		985	738

* count = number of students having that number of false alarms

As for the pseudowords themselves, false alarms range from four through 450 for the individual pseudowords, with a mean of 30.78 false alarms per pseudoword, and a SD of 77.59. Closer inspection also showed that one of the 32 pseudowords was reported much more than any other. The pseudoword "tought" received 450 (45.7%) of the total 985 false alarms, likely because of its resemblance to the past tense of teach. Perhaps reflecting the lower-level university students' limited vocabulary, they account for only 48% of the false alarms on this particular pseudoword, despite making up 58.4% participants. (For a full list of pseudowords and their reports, see Appendix).

Conclusion

This paper has been an investigation into whether or not low-level Japanese university English learners check more pseudowords in Y/N vocabulary tests than higher ability students as found in previous European research (Meara, 1996; and Cameron, 2002). A large sample (n = 738) took one of two versions of a Y/N vocabulary test which contained 40% pseudowords. The low overall false alarm rate of just over 4% in this study suggests that pseudowords were not problematic for these Japanese university students. This was further substantiated by the fact that over 70% of students had less than 2 false alarms out of a maximum of 32. This low false alarm rate suggests that these students, in general, did not claim knowledge of words they did not actually know the meaning of.

On a university-by-university basis, the two lowest-level institutions had lower false alarm rates than the two highest-level universities. A high-low analysis confirmed that the low-half of students had a 27.8% lower false alarm rate than the high-half of participants. This lower false alarm rate may be a result of greater conservatism (Huibregtse, et al., 2002) possibly resulting from lower confidence by the low-level students concerning their lexical knowledge. Alternatively, higher-level students may have greater confidence in their lexical knowledge and thus may be less cautious when facing pseudowords. The majority of false alarms on the pseudoword *tought* being from the higher-level students seems to confirm this view.

This study suffers from a number of limitations. About 9% of the test forms had to be excluded from the results either due to students writing the same Y/N test version, or misunderstanding the pseudoword test instructions. Also the 96 test items were on average too easy for many of these participants, as suggested by the Y/N test mean of 69.9%, so some ceiling effects, where a number of students knew all the words, did occur at the higher-level institutions. Further, as the sample was not a random sampling, the results cannot be generalized beyond the classes involved.

Despite these shortcomings, these findings could be of considerable interest to vocabulary researchers working with Japanese university students. That low-level Japanese students do not appear to have more difficulty with the pseudoword format should encourage test developers to include them in high-stake vocabulary tests such as placement tests, where participants have something to gain from overestimating their lexical knowledge.

Bio data

Raymond Stubbe is a lecturer at Kyushu Sangyo University. His research interests include vocabulary acquisition and testing. <raymondstubbe@gmail.com>.

Chris O'Sullivan has been teaching at Kitakyushu University for nine years. He is currently interested in trilingualism.

Jeremy Boston teaches at Tokyo University of Foreign Studies and is currently studying for his PhD in Applied Linguistics at the University of Leicester. <jeremyboston@msn.com>.

Mathew Porter is a lecturer at Hiroshima Shudo University. He is interested in CALL. <portermathew@hotmail.com>.

Richard Grumbine is associate professor of English at Ariake National College of Technology. He specializes in critical thinking and TESOL.

David Latz is a Brisbane-born lecturer at Shimonoseki University. His current research interests include pragmatics and motivation in language learning.

References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark: International Reading Association.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In Hutson, B. A. (Ed.), *Reading comprehension and the assessment and acquisition of word knowledge (Vol. 2)*. Greenwich: JAI Press

Bauman, J., & Culligan, B. (1995). *About the GSL*. Retrieved from <http://jbauman.com/gsl.html>.

Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & Van de Velde, H. (2001). Examining the yes-no vocabulary test: Some methodological issues in theory and practice. *Language Testing, 2,* 235-274.

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research, 6 (2),* 145-173.

Chall, J. S. & Dale, E. (1950). Familiarity of selected health terms. *Educational Research Bulletin, 39,* 197-206.

Chapelle, C. A. (1998). *Construct definition and validity inquiry in SLA research.* Cambridge: Cambridge University Press.

De Veaux, R., Velleman, P., & Bock, D. (2008). *Stats: Data and models* (2nd ed). Boston: Pearson Education.

Eyckmans, J. (2004). *Measuring receptive vocabulary size.* Utrecht: LOT (Landelijke Onderzoekschool Taalwetenschap).

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., & Schmitt, N. (2009). Developing an integrated diagnostic test of vocabulary size and depth. *Regional Language Centre Journal 40(1),* 5-22.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Testing, 54(3),* 339-436.

Meara, P. (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K. and Williams, J. (Eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge: Cambridge University Press.

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

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle.

Nation, I.S.P., & Belgar, D. (2007). The vocabulary size test. *The Language Teacher, 31(7),* 9-12.

Nation, I. S. P. & Cobb, T. (nd). *BNC 20.* Retrieved from <http://www.lextutor.ca/vp/bnc/>.

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.

Stubbe, R. (2010). Exploring the lexical challenge of the TOEIC® Bridge. In A. M. Stoke (Ed.), *JALT 2009 Conference Proceedings.* Tokyo: JALT.

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.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30,* 79-95.

West, M. (1953). *A general service list of English words.* London: Longmans Green.

Appendix 1

List of 32 pseudowords and their number of false alarms (n = 732)

pseudoword	false alarms		pseudoword	false alarms
tought	450		troit	13
elactroc	58		wirt	13
thuns	53		prufar	12
masuc	42		boash	11
torm	36		hebbary	11
cuttem	29		plocious	11
iligan	21		gellack	10
lowmar	20		haightid	10
ofaciel	20		maintome	10
fimul	19		thraegh	9
ebide	17		haxt	8
sweck	15		muchonasm	8
knoo	14		veye	8
gonarel	13		narsaby	7
spoy	13		sliw	7
thait	13		siccood	4