Establishing an Online Vocabulary Levels Test by Using Flash Incorporating COPS Theory

Kazumi Aizawa[1] and Tatsuo Iso[2]

Abstract. The present study aims to demonstrate how the estimation of vocabulary size might be affected by two neglected factors in vocabulary size tests. The first factor is randomization of question sequence, as opposed to the traditional high-to-low frequency sequencing. The second factor is learners' confidence in choosing the correct meaning for a given target word. A new online vocabulary size test was developed for the purpose of the study with the two factors in mind. The results of the test revealed that (1) randomizing question sequences did not have significant effects on the score of the vocabulary size test and (2) even though the learners who had a mastery level of 8000 words showed higher confidence in high frequency words than the learners with a smaller vocabulary, such confidence faded as early as 4000 frequency level of JACET 8000. The findings are discussed in detail in terms of the scale or the length of vocabulary size tests as well as the need for incorporating confidence in the estimation of vocabulary size.

Keywords: vocabulary size, multiple-choice test, confidence level.

1. Introduction

It is widely accepted that the knowledge of vocabulary is one of the most important and fundamental assets one would hope to attain in order to carry out a task involving verbal communication more successfully. Accordingly, many attempts have been made to measure the outcome of vocabulary learning.

1. Tokyo Denki University, Tokyo, Japan; aizawa@cck.dendai.ac.jp

2. Reitaku University, Chiba, Japan

How to cite this article: Aizawa, K., & Iso, T. (2013). Establishing an Online Vocabulary Levels Test by Using Flash Incorporating COPS Theory. In L. Bradley & S. Thouësny (Eds.), *20 Years of EUROCALL: Learning from the Past, Looking to the Future. Proceedings of the 2013 EUROCALL Conference, Évora, Portugal* (pp. 1-1). Dublin/Voillans:

Such attempts yielded vocabulary tests of many kinds that are appreciated and enjoyed among teachers and researchers who recognize the importance of vocabulary and wish to have deeper insights into the nature of vocabulary and its growth. Among such tests, vocabulary size tests have received most attention so far. A few examples would be Vocabulary Levels Test (VLT) (Nation, 1990) and Yes/No Test (Meara, 1992).

Despite their popularity, however, there are few studies conducted on the limitations of the vocabulary size tests (Aizawa, 2006a, 2006b; Aizawa & Iso, 2007). Although it has been shown that the test scores from which the learner's vocabulary size is estimated vary depending on the types of vocabulary tests, we have yet to see how several factors of a vocabulary size test could affect the results. One such factor is the sequence of questions. We believe this is an important issue when the time required to complete a vocabulary size test becomes longer, since learners can become more susceptible to fatigue in the latter part of the test.

Confidence is also a factor that has not been paid attention to. Researchers and practitioners intuitively know that learners do not necessarily answer questions with the same degree of confidence when taking a vocabulary test, especially when it is a multiple-choice test. Some questions will be answered highly confidently while others with lower confidence or with no confidence at all when guesswork is employed. What we do not know yet is how the concept of confidence can be incorporated in the design of vocabulary size tests by means of Clustered Objective Probability Scoring (COPS) (Shizuka, 2004), for example. The present study, therefore, discusses how such vocabulary test factors might affect the estimation of learners' vocabulary size.

2. Study

2.1. Purpose

The current study primarily aims to investigate how the ordering of questions affects the estimate of learners' vocabulary size. It also attempts to include the measurement of learner confidence in a vocabulary test and investigates the relationship between the estimated vocabulary size and learners' confidence in answering each question of a vocabulary test. Research questions are as follows.

- What are the effects of randomizing the order of questions in a vocabulary size test?

- How does confidence interact with learners' vocabulary levels as well as the frequency levels of vocabulary?

2.2. Participants

A total of 159 Japanese learners of English from two universities participated in this study. Among them, 65 subjects came from one university where they majored in English. It was expected that their overall English proficiency was slightly higher than the rest of the subjects, 94 to be exact, who were technology majors from another university.

2.3. Instrument

The Flash VLT is a multiple-choice type of test that measures learners' receptive vocabulary size (cf. for example, Schmitt, Schmitt, & Clapham, 2001). A set of three question items is displayed at the upper side of the screen. To answer, a test taker simply drags the solid circle attached to an English word and drops it to fill one of the small holes directly below the corresponding Japanese word. A hole marked with a double circle should be filled if a test taker is 90-100% confident that s/he chose the correct answer. Likewise, a hole with a single circle indicates medium confidence and one with a triangle shows that s/he does not have confidence at all.

The test adopted the target words from JACET 8000 (JACET, 2003). The list is divided into eight levels based mostly on frequencies, with each level containing a thousand words. From each level, 30 words were randomly chosen as question items. During the selection of the items, an effort was made to keep the ratio of the part of speech to as close as that of the original subsists so that the question items are the better representatives. The total number of question items is 240 (30 words x 8 levels).

Two slightly different versions of the same test were prepared for the purpose of the study: FIXED and RANDOM. In the FIXED version, the 80 sets of three target words are in descending order of word frequency. The RANDOM version only differed from the FIXED version in the sequence of the question items. Each time the test started, the same 80 sets of target words were automatically randomly sequenced except the first three sets. The order of the first three sets were fixed in order to identify the subjects who did not understand the directions of the test and failed to choose correct answers to the target words that they most likely have already learned before.

2.4. Procedure

All the subjects took both versions of the test with exactly one week in between. Half of the subjects took the FIXED version first, and then took the RANDOM version. The order of the two versions was reversed for the other half of the subjects. All of the subjects finished each version of the test within 40 minutes.

3. Results and discussion

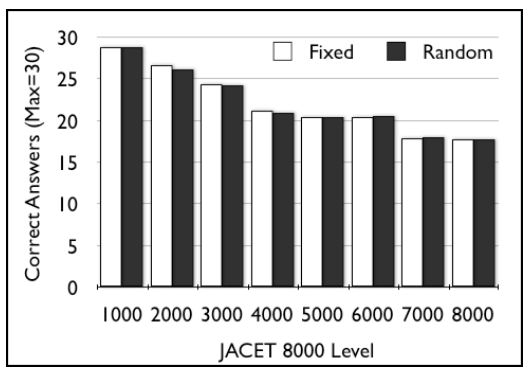
The scores of the two versions of the Flash VLT was compared to find out if randomizing the question sequence of a vocabulary size test would yield different outcomes when compared to the traditional "higher-to-lower frequency" order. The results showed that the estimated vocabulary sizes obtained from the two versions of the same test did not statistically differ (see Table 1 and Figure 1). Moreover, when the test results were examined by each frequency band, it was apparent that the subjects performed in the same manner in the two versions. Considering that the subjects were to repeat the form-meaning matching task more than 200 times, it was expected that fatigue would negatively affect the subjects' performance in the FIXED version, especially since the words with lower frequency were arranged toward the end of the test.

Table 1. Descriptive statistics of the vocabulary size (N = 159)

	Mean*	SD
FIXED	5924.2	949.5
RANDOM	5907.0	949.1

* The maximum possible vocabulary size was 8000.

Figure 1. Comparison between FIXED and RANDOM (N = 159)



As for the relationship between confidence and learners' vocabulary levels as well as word frequency, the overall results were generally in accordance with our expectation. The larger the vocabulary the subjects acquired, the more confident they were. Also, the less frequent the target words became, the less subjects showed confidence. The results indicate that flash VLT successfully elicited learners' judgments on their own confidence.

On closer examination (Table 2), there were clear patterns in the decline of confidence across the vocabulary level groups. The subjects in groups below 4000 vocabulary level lost their confidence as early as 2000 level target words, whereas those in G4000, G5000, and G6000 maintained their confidence that they had when dealing with 1000 level words at least until the end of 2000 level target words. Further, the groups with the highest vocabulary levels, G7000 and G8000 continued to be as confident until 3000 level words as they were dealing with 1000 level words. What can be inferred from here is that obtaining a passing grade of 80% at a certain level frequency level in vocabulary size test does not necessarily ensure that learners are dealing with questions with high confidence. The question is what it means to have a 6000 vocabulary level when such learners are not very confident in dealing with 4000 level words. Apparently, the concept of vocabulary level (and vocabulary size as well) needs to be reconsidered with confidence in mind if learners' size of vocabulary is to be quantified.

Table 2. Distribution of the answers with "high" confidence (%)

		JACET 8000 Levels							
	N	1000	2000	3000	4000	5000	6000	7000	8000
G1000	25	85	59	53	42	44	42	34	30
G2000	27	92	77	62	53	47	49	39	38
G3000	48	97	86	78	59	56	58	40	43
G4000	23	97	90	82	72	61	65	52	52
G5000	11	98	94	87	72	69	67	53	59
G6000	13	99	96	89	75	70	69	48	54
G7000	5	100	98	94	80	73	77	63	63
G8000	4	100	98	94	90	90	90	84	85

4. Conclusions

The findings of this study confirmed the traditional testing methodology of receptive vocabulary size in terms of how the question items should be sequenced. They also demonstrated how confidence should be taken into account when estimating the size of vocabulary. For further research, it will be of high importance to investigate

how to incorporate learners' confidence in the calculation of vocabulary size estimations through vocabulary size tests.

References

Aizawa, K. (2006a). Rethinking Frequency Markers for English-Japanese Dictionaries. In M. Murata, K. Minamiide, Y. Tono, & S. Ishikawa (Eds.), *English Lexicography in Japan*. Tokyo: Taishukan.

Aizawa, K. (2006b). Is Frequency of Words Related to Difficulty of Words? *Paper presented at the BAAL/IRAAL*.

Aizawa, K., & Iso, T. (2007). Estimating Word Difficulty: The Divergence From Frequency Levels. *Annual Review of English Language Education in Japan (ARELE), 18*, 111-120. Retrieved from http://ci.nii.ac.jp/lognavi?name=nels&lang=en&type=pdf&id=ART0009707101

JACET. (2003). *JACET List of 8000 Basic Words*. Tokyo: The Japan Association of College English Teachers.

Meara, P. (1992). *EFL vocabulary tests*. University of Wales, Swansea: Center for Applied Language Studies.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of the two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88. doi: 10.1177/026553220101800103

Shizuka, T. (2004). *New horizons in computerized testing of reading*. Osaka: Kansai University Press.