

Second Language Vocabulary Assessment

Yuji Nakamura¹, Adam Murray² and Haruhiko Mitsunaga³

¹ Keio University, ²Tokai University, ³Tokyo Institute of Technology

nkyj@flet.keio.ac.jp, murray74@nifty.com, mitunaga.h.aa@m.titech.ac.jp

Abstract

Vocabulary knowledge is an important factor in measuring reading skill in EFL. In this study, the validity of the vocabulary test that has been administered at Keio University Faculty of Letters since 2006 was examined through 1) comparing existing readability scores, and 2) counting how often JACET8000 (JWL) (JACET, 2003) low-level and high-level words appeared as the options to test items. We found a negative correlation between readability and item difficulty. Further analysis suggested that there were high JWL level options and low JWL level options regardless of item difficulty, and there was no evidence that high-level JWL words, even off-list words, were the most enticing options. As a result, we conclude that both readability and JWL word level cannot be considered an indicator of item difficulty in vocabulary test for the purpose of measuring reading skill.

Keywords

JACET8000, vocabulary measurement, construct validity, placement test

Background

It can be argued that vocabulary is essential to measuring proficient reading levels in both first and second languages. While there has been extensive discussion about the importance of teaching vocabulary in academic settings, testing vocabulary itself has received less attention. Currently, the validity of vocabulary testing as a distinct construct is a matter of debate, whereas the assessment of reading and reading comprehension has consistently been linked to some form of vocabulary measurement (Ekbatani, 2011).

In general, there are two approaches to define 'difficulty' or how difficult it is for EFL learners to acquire vocabulary. One method is based on the 'word level', which is defined as the requirement of reading English. The other is based on a testing approach. The standardized test using item response theory (IRT) enables us to compare the difficulty of items in multiple placement test forms, and to

construct a common scale of difficulty.

1 Purpose of this study and research questions

The purpose of the present study is twofold: 1) to discuss important issues in second language vocabulary assessment in academic settings, and 2) to analyze the vocabulary section of a university placement test.

We set the following three research questions. First, we set research question (RQ)1 that is essential to constructing test items: What is the relationship between the word level of distractors and their attractiveness? Does the relationship change by item difficulty? Next, there are multiple words that do not appear in the JACET8000 (JACET, 2003) basic word list which are widely used in EFL. Test items with high levels of difficulty on an IRT scale often contain options with high word levels and some of the items that contain off-list words are ranked as being more difficult. To examine this hypothesis we set RQ2: Are off-list words the most enticing options among all item difficulty levels? If so, the attractiveness of the distractor does not correspond with the difficulty. Finally, the correspondence between item difficulty and word level of options can be investigated. However, item readability is not controlled in this placement test. We set RQ3: Are readability and word level good indicators of the difficulty of vocabulary test items? To address this RQ, we adopted the Gunning-Fog index score (G-F) as a readability scale and JACET8000 (JWL).

2 Method

2.1 Subjects

Since 2006, Keio University Faculty of Letters has administered an in-house placement test. Students are given placement tests twice, once at the beginning of the academic year and again at the end of the semester. For this analysis, the results of 10 exams (approximately 800 test-takers each) are examined.

Table 1: Distribution of Items by Difficulty in Each JW L

| Difficulty | L1+ | L2 | L3 | L4 | L5 | L6 | L7 | L8 | Off |
|------------|------|------|------|------|------|------|------|------|------|
| Easy | 32.7 | 26.7 | 36.4 | 42.5 | 48.8 | 47.1 | 48.8 | 34.1 | 32.8 |
| | 33.4 | 34.5 | 40.5 | 48.7 | 51.4 | | 54.8 | 43.8 | 37.3 |
| | 37.3 | 34.7 | 40.7 | 52.7 | 51.8 | | | 53.7 | 39.3 |
| | 38.9 | 37.8 | 43.0 | 56.2 | 52.8 | | | | 52.6 |
| | 39.7 | 42.1 | 47.6 | | 70.7 | | | | 64.6 |
| | 40.2 | 50.4 | 54.2 | | | | | | 66.5 |
| | 40.5 | 56.0 | 57.4 | | | | | | 69.0 |
| | 41.1 | 75.1 | 59.0 | | | | | | |
| | 48.9 | | 67.6 | | | | | | |
| | 56.5 | | | | | | | | |
| Difficult | 56.9 | | | | | | | | |

Note: Excludes overlapping words.

2.2 Materials

The placement test is a 60-minute examination which consists of 50 questions in four sections: grammar, vocabulary, gap-fill (cloze), and reading. For this study, only the vocabulary section was examined. The vocabulary section consists of 10 multiple choice questions with four options. The contents of 10 placement tests, and 100 test items were examined.

2.3 Analyses

First, we calculated the percentage of JW L for each of the options (Analysis 1). Next, the item difficulty scale by IRT was constructed using WINSTEPS (Linacre, 2005), and the items were ordered by the IRT difficulty scale. After this, we excluded overlapping alternative words (Analysis 2). We calculated the proportion of selection for each option by JW L to answer RQ1 (Analysis 3). In Analysis 3, we also excluded items that more than 40% of the students responded correctly because our main focus was the items that tend to get answered incorrectly. Finally, we examined the attributes of test items by item discrimination and difficulty, and calculated G-F (Analysis 4).

3 Results and Discussion

From Analysis 1, we found that Level 1 and Plus words were the most frequently used (27%), followed by Off-list words (18%), and Level 6,7,8 (5%) were the least frequently used.

Table 1 indicates that words in each JW L have various levels of difficulty as a result of Analysis 2. From Table 1, even Off-list words often appeared as both easy and difficult items, whereas Level 1 words appeared frequently regardless of their item difficulty. Thus, the answer to RQ2 seems that the difficulty of items has little to do with the word level of the options. As a result of Analysis 3, 30 items were selected, and we found both high and low level words which appeared as distractors that

were more or less attractive. From the answer to RQ2, item difficulty has little in common with the word level, then the answer to RQ1 is as follows: it seems there is a weak relationship between word level and attractiveness for any level of item difficulty. Table 2 indicates the correlations between the pair of JW L, G-F, item difficulty and item discrimination. There is a weak negative correlation between readability and item difficulty, and a weak negative correlation between JW L and item difficulty. As a result, we conclude that neither G-F nor JW L can be considered good indicators of difficulty.

In this study we used the placement test data for assessing the reading skill of examinees. If this test measures a different aspect of vocabulary skill from JW L, it is natural that G-F and JW L do not perform well. Further study is needed to investigate construct validity of EFL vocabulary tests.

Table 2: Correlations between JW L, G-F, Item Difficulty (DIFF.) and Discrimination (DISC.)

| | Level | G-F | DIFF. | DISC. |
|-------|-------|-------|---------|--------|
| Level | 1 | .198* | -0.271* | -0.028 |
| G-F | | 1 | -0.181 | 0.088 |
| DIFF. | | | 1 | 0.594* |
| DISC. | | | | 1 |

Note: * $p < .05$.

Acknowledgement

This study was supported by the Grant-in-Aid for Scientific Research (KAKENHI) (C) (24520646).

References

- Ekbatani, G. (2011). *Measurement and evaluation in post-secondary ESL*. NY: Routledge.
- Japan Association of College English Teachers. (2003). *JACET list of 8000 basic words*. Tokyo: JACET.
- Linacre, L. (2005). *WINSTEPS Version 3.55* [Computer Program]