

# Multiple Choice Tests and Distractor Analysis in the Japanese University Classroom

Richard Derrah<sup>1</sup> and D.E. Rowe<sup>2</sup>

<sup>1</sup>Kindai University, Japan

<sup>2</sup>Kobe Women's University, Japan

**Abstract:** *This paper describes the results of a study investigating the use of a multiple choice vocabulary size test in the Japanese university setting. The vocabulary test was administered to students (N=520) in 2017. This study investigates the appropriateness of this test for these students. Distractor analysis was also included in order to improve the test for future administrations. Distractor analysis examines the functioning of incorrect answers within the test. Attendees to this presentation will become familiar with the use of the vocabulary size test used in the study and also develop an understanding of the benefits of conducting distractor analysis for multiple-choice tests.*

**Keywords:** *vocabulary, distractor, Rasch analysis*

## 1. Introduction

In the language classroom multiple-choice tests provide a useful tool to assess student performance. They are often applicable to a wide variety of purposes. However, writing and developing multiple-choice tests can take a substantial investment of time. Part of the challenge is to develop attractive distractors (incorrect choices) for test items, which may be attractive for test takers, but still not the best choice for a particular item. The effectiveness of a test is often based upon the performance of the students taking the test. Another way to view the effectiveness of a test is to evaluate how students interact with the distractors. This paper details one method to examine the effectiveness of distractors on a multiple-choice test. The example used here is a multiple-choice test of English vocabulary size designed for native speakers of Japanese who are studying English as a second language.

## 2. Methods

### 2.1. The Vocabulary Size Test

Vocabulary is a primary component of learning a new language. For learners of English the time spent on the acquisition of vocabulary is a vital part of their fluency development. For teachers, having an understanding of the approximate size of a student's vocabulary is beneficial to develop appropriate goals and objectives. The vocabulary size test as developed by Paul Nation is designed to provide an approximation of non-native English speakers' vocabulary size. Words are divided into bands based upon frequency as determined from corpus studies. Each band represents 1,000 words of English based upon frequency. The vocabulary size test is a multiple-choice test with ten questions for each band of word frequency. In other words, each set of ten questions represents a band of 1,000 words. The first ten questions are words representative of the most frequent 1,000 words of English. The second ten questions thus represent the second most frequent 1,000 words of English. The test used in this study included words covering the most 5,000 most frequent words of English.

The test has multiple different versions of including several bilingual versions available on Paul Nation's website (<http://www.victoria.ac.nz/lals/about/staff/paul-nation>). Research on some of these versions of the vocabulary size test including the Iranian-bilingual version (Karami, 2012), Vietnamese-bilingual version (Nguyen & Nation, 2011), and the Russian-bilingual version (Elgort, 2013) have shown that these versions are more accurate in measuring vocabulary size than versions which only use English.

Each item on the test has a prompt including the target vocabulary item and the word used in a simple sentence. The sentences are designed not to provide a hint for meaning. The prompt is followed by four answer choices. Only one answer is possible for each question. An example question from the test used in this study is below:

Time: They have a lot of time.

- 1) お金
- 2) 食べ物
- 3) 時間
- 4) 友だち

The three distractors are お金 [okane (money)], 食べ物 [tabemono (food)], and 友だち [tomodachi (friend)]. The correct answer is number three: 時間 [jikan (time)]. Some versions of the vocabulary size test use English for both the item prompt and possible answers. Some researchers such as .... Have shown that providing question answers and distractors in the native language of the test taker has higher reliability. If English is used for both elements of a test item then the students are, in effect, being tested twice since in order to respond correctly the students would need to understand both the test prompt and the distractors.

The test was given to the entire freshman class of the applied sociology department of a large university in western. All of the students in this department are required to take two classes of English during their freshman year. One conducted generally conducted by a native speaker of English focusing on listening and speaking with the other class provided by a native speaker of Japanese focusing on reading, writing, and grammar. The website for the test was given to the students on a handout in class. The handout included the webpage address and a QR code for the address. The QR code allowed students to directly access the test from their smartphones or other devices capable of scanning QR codes. This allowed students to access the test at their convenience. In addition, instructions on the test were provided in Japanese.

## 2.2. Rasch Analysis

Rasch analysis is a dichotomous model of assessment. In Rasch analysis both item difficulty and the test takers' ability are calibrated into equally distant units for measurement. These units are log-odd units or logits. These units allow for a better comparison of test takers' ability than simple raw scores. In logits the average ability of all the test takers on a particular test is set at zero. Students with logits of ability higher than zero are about the mean while those below zero are below the average ability of that group of test takers. The same is true for item difficulty, which is also calculated for each item on a test in logits. For this study, Winsteps software (Linacre, 2010) version 3.69.1.15 (<http://www.winsteps.com/index.htm>) was used to conduct Rasch analysis.

## 2.3. Distractor Analysis

Distractor analysis allows researchers (and others) to examine how the distractors on a multiple-choice test are interacting with the test takers. By examining distractors, test makers can see which distractors are attractive for test takers. Items, which are not selected by any test takers, are obviously wrong and not an attractive option. This type of distractor should be rewritten to provide a more challenging test. In addition, the average logit of ability of test takers is shown for each answer choice. With this feature it is possible to see if test-takers with higher average logits of ability are selecting wrong answers when others with lower average ability are selecting the correct answer. In other words, higher ability students are getting the question wrong when comparably

lower ability test-takers are answering the questions correct. This situation shows that the question is not performing as it should and is in need of investigation.

### 3. Results

The results of the distractor analysis are shown below in table one. In this table the entry number indicates the item number on the test. The data code is a list of the four multiple-choice options for each test item. In addition, missing accounts for students who did not answer the question. Score value is marked as a zero for an incorrect choice and one for the correct answer. The count column displays the number of test-takers who selected each distractor or correct answer choice. The average measure is the average logit if ability of all of the test-takers who selected each of the options for each test item.

Table one shows the distractor analysis for a representative three items out of the fifty on the vocabulary size test used in the study. For these items the distractor analysis shows that item number 16 is the most problematic out of the group. For item number 16, the average logit of ability of the test-takers choosing the correct answer (option 2) was lower than the average logit of ability of those choosing options numbers 4 and 1. For this item, students who performed better on the test selected an incorrect answer. This is an indication that this item is, for some reason, causing a problem for students with higher logits of ability and should be investigated further and possibly be rewritten.

Item numbers 45 and 34 do not have the same issue as the test-takers with higher ability selected the correct answer on average. For item number 45, the average logit of ability for students selecting the correct answer was 65.30 and the average ability of test takers for the distractors were all lower. Item number 34 showed the same results indicating that this item is working well for this particular group of test-takers. In addition, this analysis also shows that all three of these questions are rather difficult as only about 10% of test takers selected the correct answer.

TABLE I: Displacement Order

Entry Number	Data Code	Score Value	Count	%	Average Measure
16	3	0	52	10	59.44
	4	0	83	16	62.39
	1	0	342	68	63.65
	2	1	29	6	61.77*
	missing	***	11	2	68.57
45	3	0	11	2	57.92
	2	0	148	29	62.12
	1	0	293	58	63.11
	4	1	53	10	65.30
	missing	***	12	2	65.04
34	3	0	32	6	58.23
	4	0	192	38	62.71
	1	0	218	43	63.38
	2	1	64	13	64.24
	missing	***	11	2	68.01

### 4. Discussion

Distractor analysis can help teachers and researchers investigate the functioning of multiple-choice tests. Considering the considerable investment of time and resources in test creation, distractor analysis provides a tool to improve tests. This has two aspects. First, items, which are more difficult for higher ability students, are identified and flagged for further investigation. This could indicate something such as higher ability students overthinking an item causing them to get the question wrong. It is important for teachers and researchers to be

aware of that issue within the test. Second, distractor analysis can show if some distractors are not being selected by test-takers. If no one selects a distractor that indicates that the distractor is too easy and instead of that item function as a four choices, effectively students are selecting from only three items. For these reasons distractor analysis is an effective tool for language teachers.

## 5. References

- [1] Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, 30(2), 253-272.
- [2] Karami, H. (2012). The Development and Validation of a Bilingual Version of the Vocabulary Size Test. *RELC Journal: A Journal of Language Teaching and Research in Southeast Asia*, 43(1), 53-67.
- [3] Linacre, J. M. (2010). A User`s Guide to WINSTEPS. Chicago: winsteps.com
- [4] Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*(31), 9-13.
- [5] Nguyen, L. T. C., & Nation, P. (2011). A Bilingual Vocabulary Size Test of English for Vietnamese Learners. *RELC Journal: A Journal of Language Teaching and Research in Southeast Asia*, 42(1), 86-99.