

# Validating the Japanese Bilingual Version of the Vocabulary Size Test

Richard Derrah and D. Edward Rowe

**Abstract**—This study uses Rasch analysis to validate the Japanese bilingual version of the 14,000-word level vocabulary size test. The test was given to two sets of students. The first group were students at a women's university in Western Japan (N=27). The second group consisted of students from a high school also in western Japan (N=16). All participants (total N=43) completed the 140 question Japanese bilingual version of the 14,000-word level Vocabulary Size Test. The data was then analyzed using Winsteps software to perform Rasch analysis to check the validity of the test.

**Index Terms**—Vocabulary, Rasch analysis, testing.

## I. INTRODUCTION

Vocabulary is a central part of language learning. Nation [1] divides language learning into four equal parts or strands: meaning-focused input and output, language-focused learning and fluency development. They can be implemented to help a second language learner in all aspects of language learning with vocabulary an important characteristic in all four strands. Meaning-focused input includes learning while listening or speaking [1] in which the learner has a 98% of the running words. The higher a learner's comprehension level of vocabulary the more they can benefit. Language-focused learning involves studying specific aspects of the language like vocabulary. Thus the study of vocabulary in the language-focused learning can benefit the other strands including meaning-focused input. The remaining two strands: meaning-focused output and fluency development are more successful both in study and practicality with a high or improving comprehension of vocabulary. Nation [1] describes simplified texts or easy readers as a method for students to improve their level of vocabulary comprehension.

Barcroft [2] also described vocabulary as a key for SLA students in 3 aspects: their ability to communicate, their self-realization in the importance of vocabulary, and in their use in grammar. Barcroft [2] also gives five effective principles for student's learning vocabulary in SLA. His first principle includes the presentation of new and a repetition of difficult words either by or a combination of direct or indirect instruction. Bancroft's [2] principles also include the progressive instruction of vocabulary from easy to difficult and connects the principles in relation to input and input processing. Green [2] also connects the amount of information we retain with the number of times we study it.

Manuscript received October 3, 2014; revised April 4, 2015.

Richard Derrah is with the Kwansai Gakuin University, Japan (e-mail: rderrah@yahoo.com).

D. Edward Rowe is with the Temple University Osaka, Kobe Women's University, Japan (e-mail: dethrowe@gmail.com).

The more opportunities a student has learning a word or words the more likely they are going to retain it.

Being able to measure a student's vocabulary size with a valid instrument can be a benefit to students, teachers, administrators, and researchers. Using lists from the British National Corpus (BNC) as a base, the Vocabulary Size Test (VST) [1] was developed to measure vocabulary size knowledge up to the 14,000 word level. Based upon their frequency, vocabulary items are placed in bands of 1,000 words. The VST has 140 questions in total made up of 10 multiple choice questions per 1,000-word level. Each question has four multiple-choice options. On the original test these options were in English, however, on the bilingual versions only the question prompt is in English while the four answer choices are in the native language of the projected test taker. For example, the Japanese bilingual test has the answer options in Japanese. An example of a test item from the Japanese bilingual test is below:

Shoe: Where is your shoe?

- 1) 親
- 2) 財布
- 3) へん
- 4) くつ

There are several bilingual versions of the VST available on the Paul Nation's website: (<http://www.victoria.ac.nz/lals/about/staff/paul-nation>).

Research on some of these versions of the VST including the Vietnamese-bilingual version [2], Iranian-bilingual version [3] and the Russian-bilingual version [4] indicate the bilingual versions of the VST are a more accurate measurement of a student's vocabulary size than the English-only version of the VST. If the answer choices are in English, the test-taker needs to understand the test prompt as well as all of the answer choice options in contrast to the bilingual versions that provide students answer options in their native language. This means the test is focusing on the test items rather than having the possibility of the student understanding the test prompt in English, but unable to understand the options provided in English.

## II. PARTICIPANTS

The participants in this study consisted of groups of students from two different schools. Each group will be described in detail in this section. The first group (N=27) attends a women's university in Western Japan. All are third year students and attend either a Communication class, ToEIC/TOEFL class, or both. All of the Communication's class students have a major that is linked to English with most in English literature or English education. Some had had

homestay experience in the United States. The Toeic/Toefl class students are a more diverse mix between English and non-English majors. Both classes are 90 minutes long and held once a week for fifteen weeks. The students have a median Toeic score of approximately 500.

The Communication's Class is required for all English related majors. Class time was divided between listening, speaking, vocabulary, and focus on form. There were 29 students in the class so a lot of class time is used with pairs and small groups. The base textbook was the everyday English conversation textbook *English Firsthand*. Most classroom activities were tied in, or expanded from, if not directly from the textbook. The exercises included: forming questions about countries, student's family and history, personal experiences, descriptions, and directions. The students would often break-up into small groups and exchange stories with others in their groups. A few members from each group would do their exercise in front of the class. This was rotated between group members over the term. Students had weekly homework assignments usually asking them to connect a personal story or experience from a subject discussed in class and some grammar exercises. The vocabulary was thus connected to everyday English. Words from recent exercises included: heavysset, stocky, interrupts, and balding.

The Toeic/Toefl class is focused on the students improving their Toeic/Toefl score and is held in a classroom that gives every student access to a computer. All students had taken either the TOEIC or TOEFL test and were focused on improving their score. There were 25 students in the Toeic/Toefl class. The students had three mini Toeic/Toefl tests, one regular Toeic test, one step-test, and weekly self-study exercises. The students also had self-study exercises they could individually practice by computer. A typical class was divided in half. Students would practice various Toeic/Toefl exercises during half of the class and the other half they would do self-study exercises. The exercises were rotated around the different parts of the Toeic/Toefl tests: listening, grammar, reading, writing Toefl, and speaking Toefl.

However, though most exercises focused on vocabulary the student's often had had little knowledge and no experience with the words from the practice section of the test. Especially, the reading sections of both the TOEIC and TOEFL sections require extensive vocabulary study. The test reading section questions had many vocabulary words and phrases that the students were unfamiliar and well under the 98% Nation's gives as effective for meaning-based input. But follow up exercises were a mix of common and test related vocabulary words. Homework consisted of vocabulary and grammar exercises based around the TOEIC/TOEFL test exercises. Most students kept a notebook with a list of the vocabulary words discussed in the class exercises or covered in the homework. Vocabulary words of a recent homework assignment which were appropriated from practice TOEFL tests included: *inquiry, marshaled, diffuse, balmy, poured over, intersection, and repugnant* [Appendix A]. As Barcroft [2] noted in his principles, repetition of difficult words is key for a student learning vocabulary. Barcroft [2] also talks about the importance of learning vocabulary root words which another

common type of class homework assignment.

Barcroft [2] also discusses the importance of allowing students the students to comprehend new or unfamiliar vocabulary without forcing the student to use the new words in output. Students in both classes had several opportunities to comprehend and intake the vocabulary before using the vocabulary in output.

The second group of participants included students from a private high school in Western Japan. The students have several hours of English instruction per week conducted in Japanese with 1 hour of a general English communication class with a native speaker of English. Most of the English classes are conducted with the purpose of helping the students to do well in their upcoming university entrance examinations. These examinations are extremely competitive and are mostly focused on English grammar rather than communication. The students were all second year high school students ( $n = 16$ ).

Activities for the classes conducted by native speakers of Japanese focused on reading and grammar translation based activities. Translations were from English into Japanese for selected example sentences within a reading. Some vocabulary instruction was conducted, but this was only to support translation activities rather than to expand vocabulary knowledge.

In the English communication classes conducted by a native speaker of English, activities were focused on preparing students for in-class presentations on topics such as education, history, and science. Some vocabulary instruction was conducted based on the topic of the presentation (such as history) in order to provide a common basis of knowledge for the students within the class. No other general vocabulary test was conducted within the course.

### III. METHODS

The 14,000 word-level bilingual Japanese version of the vocabulary size test was placed on a password protected website. An explanation of the study was then made to intact classes at both research sites and students were invited to participate on a voluntary basis. The website information including the password were distributed to the participants. A window of one month was allotted for participants to complete the test. At the end of the month the data was collected from the website.

The instructions the students received after going to the website is shown below in Fig. 1 below. Since all of the participants were native speakers of Japanese the instructions were provided in Japanese. The instructions in Japanese in Fig. 1 state "2014 English test. This test is not connected to your grade. Please do not consult with others, use the Internet or dictionaries to help with the test. Please do your best on your own."

Test data was collected from the two groups of participants and analyzed with the Rasch model by using Winsteps software [5]-[7] version 3.69.1.15 (<http://www.winsteps.com/index.htm>). The Rasch model produces calibrated scales of both the ability of the test-takers and the difficulty of the test items. In other words, the units used by the Rasch model, logits, are equally distant

unlike raw scores. In the case of raw scores, a person scoring an 80% on a test would not be considered as having twice the ability as a person scoring a 40%. In the Rasch model, however, a person with a logit of ability of 2.0 is considered twice as able as a person with a logit of ability of 1.00. This is true for test item difficulty as well. A test item with a logit of difficulty of 3.6 is considered three times as difficult as an item with a logit of difficulty of 1.2.

2014年 英語テスト

このテストは、成績に関係ありませんが、他の人と相談しないで、インターネットや辞書などを使わないで、ひとりの力で頑張ってください。

8月30日までしてください。

Fig. 1. Japanese language for students before the vocabulary test. It states: 2014 English Test. This test is not related to your course grade. Please do not consult with others, use the internet, or a dictionary. Complete the test by yourself and do your best. Please complete the test by August 30.

The use of the Rasch model will allow us to create calibrated measures of both person ability and item difficulty. These calibrated measures will then be compared to see how well the test difficulty matches the ability of the test population.

#### IV. RESULTS

Rasch analysis showed the range of item difficulty encompassed the ability of the student population taking the test. There were some items which were easier than the ability of the lower end of student ability as well as items much more difficult than the ability of the students on the high end. Having some items that are too easy and some that are too difficult for the test-taking population means the entire population is measured. The mean of item logit of difficulty and the mean of student logit of ability were very close. This indicates this test is a good match for this population.

Fig. 2 the item Wright map shows a visual representation of the match between the difficulty of the items and the abilities of the participants of this study. Each 'X' on the left side of the figure represents a number of participants. There is also an M, S, and T which represents the mean, single standard deviation, and two standard deviations respectively of the participants abilities as measured in logits. Each number on the right side of the figure represents one item on the vocabulary size test. For example, I0022 is the twenty-second item on the test. Just as on the left side of the figure there is also an M, S, and T on the right side representing the mean, single standard deviation, and two standard deviations respectively of item difficulty as measured in logits.

Rasch analysis also provides a measure of the reliability of the test. Item reliability is an indicator how well the items on this test will interact with other similar populations. In other words, the items should retain their relative ranking of difficulty. Difficult items should remain difficult while easier items should be easier for similar test takers. Item reliability was high at .92 that means the test would function in a similar fashion if given to an equivalent population. Fig. 3 shows a summary of the item measures including total score, count, measure, and fit statistics. The total score is the number of correct answers for the items. The mean of 22

indicates that, on average 22 out of 43 participants answered items correctly. The maximum number of correctly answering participants for a single item was 42 while the lowest was 3. The measure is the mean, standard deviation, maximum and minimum scores in logits for items. Finally the infit and outfit measurements are an indication of how well items fit the Rasch model. For this study items with an infit MNSQ of larger than plus or minus 1.5 was considered misfitting.

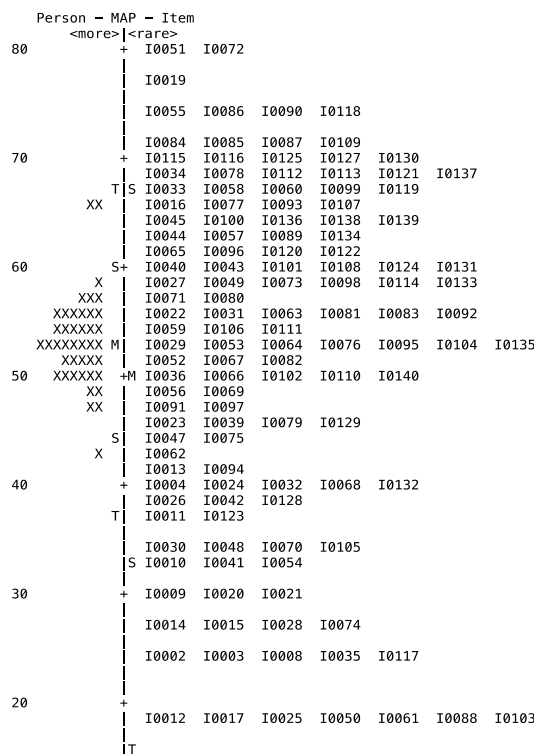


Fig. 2. Item wright map.

SUMMARY OF 134 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	22.3	43.0	50.00	4.36	.98	.1	1.01	.1
S.D.	12.1	.0	17.21	1.78	.17	.6	.64	1.0
MAX.	42.0	43.0	79.71	11.41	1.64	2.0	7.01	6.5
MIN.	3.0	43.0	8.78	3.14	.36	-1.8	.04	-9.9
REAL RMSE	4.83	TRUE SD	16.52	SEPARATION	3.42	Item RELIABILITY	.92	
MODEL RMSE	4.71	TRUE SD	16.55	SEPARATION	3.51	Item RELIABILITY	.92	
S.E. OF Item MEAN = 1.49								
MAXIMUM EXTREME SCORE:			1 Item					
MINIMUM EXTREME SCORE:			5 Item					
UMEAN=50.0000 USCALE=10.0000								

Fig. 3. Item measure.

INPUT: 43 Person 140 Item MEASURED: 43 Person 140 Item 2 CATS WINSTEPS 3.69,1.15

SUMMARY OF 43 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	74.5	140.0	52.20	2.20	1.01	.0	1.01	-.1
S.D.	13.0	.1	7.78	.30	.14	1.2	.46	1.0
MAX.	104.0	140.0	66.30	4.10	1.40	3.3	3.60	3.7
MIN.	13.0	140.0	9.53	2.13	.76	-2.6	.62	-1.9
REAL RMSE	2.30	TRUE SD	7.43	SEPARATION	3.23	Person RELIABILITY	.91	
MODEL RMSE	2.22	TRUE SD	7.46	SEPARATION	3.36	Person RELIABILITY	.92	
S.E. OF Person MEAN = 1.20								
Person RAW SCORE-TO-MEASURE CORRELATION = .98								
CRONBACH ALPHA (KR-20) Person RAW SCORE RELIABILITY = .87								

Fig. 4. Person measure.

Just as the item measures provided information on the items on the test, the Rasch model also provides these same measures for the persons within the test population. Fig. 4

person measure shows the total score, count, measure, error, and infit/outfit measures for persons. The total score shows the mean, standard deviation, minimum and maximum scores on the 140 item vocabulary test. The count is the number of questions responded to by participants. In this study all participants responded to all of the questions on the test. The measure is the mean, standard deviation, minimum and maximum scores of person vocabulary ability shown in logits. Finally, person reliability was .91 indicating this group of test-takers would respond in a similar fashion if given an equivalent test.

The spread of the items above and below the ability range of the population along with the proximity of the mean measure of item difficulty in logits (50.00) to the mean ability of persons (52.20) shows a match between the test and the target population. All of the participants with the test population were measured avoiding a ceiling or floor effect. In other words the lowest ability students were able to answer some questions correctly while the highest ability students incorrectly answered some of the more difficult items. If a student either incorrectly answered all of the questions or received a maximum score then this test would only be able to provide an estimate of the persons vocabulary size rather than a more accurate measurement which would not be an effective tool for educators.

#### V. CONCLUSION

As previous studies have indicated that other bilingual vocabulary size tests are a more accurate measure of a student's vocabulary size, the results of this study indicate that the Japanese bilingual version of the vocabulary size test is a valid test to measure the vocabulary size of Japanese English language learners. This test bounds the upper and lower limits of the population within this study providing a measure for all of the students from both the high school and university. High school and university students in Japan and beyond are a group which may most benefit from a valid vocabulary size test.

Within the Japanese context of high school and university education the 14,000 word bilingual Japanese version of the vocabulary size test is a tool to help educators measure the vocabulary size of their students. This can help to identify gaps in students' knowledge base while also helping to direct educators to vocabulary level appropriate learning materials such as graded readers. This test may also help educators to assess long-term vocabulary acquisition.

Further research on larger and more diverse groups both within Japan and without would provide more information on the strengths and weaknesses of this test. In addition, new language versions of the test could be developed to test populations beyond the currently available bilingual tests.

#### APPENDIX : TOEIC HOMEWORK EXAMPLE

##### TOEFL Vocabulary homework

1. The criminal was asked to aid police in their 'inquiry'.  
interview  
investigation  
questions
2. He 'marshaled' the troops for the offensive.

ordered

arranged

called

3. The use of penicillin is 'diffuse'.

limited

widely available

increasing

4. I like Florida because of the balmy winters.

warm

humid

cold

5. He 'poured over' the questions until he couldn't keep his eyes open anymore.

asked

examined

memorized

6. You will find the office at the main 'intersection'.

ending

crossing

roundabout

7. He found the smell of the lotus blossom 'repugnant', and asked for a different type of flower.

enjoyable

disgusting

delightful

8. Teenage 'lethargy' is often due to lack of proper nutrients.

enthusiasm

sluggishness

boredom

9. Unfortunately, information concerning the latest crisis in the state government is arriving 'piecemeal'.

incorrectly

delayed

bit by bit

10. Many of the students attitudes are considered too 'cynical' for publication.

encouraging

sarcastic

deprived

11. George Washington's 'immutable' courage was a key ingredient in the success of The United States during the Revolutionary War.

moody

upbeat

unchanging

12. It is important to weed out the 'chaff' when considering a new position.

rubbish

jewels

deals

13. His position was 'contrary' to that of the teacher's.

similar

opposite

foreseeable

14. The show 'culminated' in the presentation of the dairy princess crown.

began

ended

included

15. The 'coach' for Miami leaves shortly after 9.

- plane  
shuttle  
bus
16. Mozart's 'partiality' for chocolate from Salzburg resulted in today's famous Mozart Kugeln.  
dislike  
preference  
aversion
17. The lion's 'ferocious' behavior was attributed to its lack of edible flesh.  
tame  
relaxed  
savagely
18. Be careful not to 'rush' your studies, even when in a hurry.  
improve  
hurry  
deepen
19. Dormitory 'get-togethers' are encouraged on a monthly basis.  
study periods  
informal meetings  
cleaning sessions
20. The artist's 'profile' can be found carved into the cave walls.  
portrait  
outline  
description
21. Horses are often put out to 'pasture' during the early morning hours.  
cleaned  
field  
arena
22. Janet's 'profuse' use of the adjective like bothered the professor.  
incorrect  
delayed  
plentiful
23. The 'crooked' use of gambling schemes are often found while surfing the Internet.  
enjoyable  
difficult  
dishonest

## REFERENCES

- [1] I. S. P. Nation and J. P. Deweerdt, "A defence of simplification," *Prospect*, vol. 16, no. 3, pp. 55-67, 2001.
- [2] J. Barcroft, "Second language vocabulary acquisition: A lexical input processing approach," *Foreign Language Annals*, vol. 37, no. 2, pp. 200-207, 2004.
- [3] I. S. P. Nation and D. Beglar, "A vocabulary size test," *The Language Teacher*, vol. 31, pp. 9-13, 2007.
- [4] L. T. C. Nguyen and P. Nation, "A Bilingual Vocabulary Size Test of English for Vietnamese Learners," *RELC Journal: A Journal of Language Teaching and Research in Southeast Asia*, vol. 42, no. 1, pp. 86-99, 2011.
- [5] H. Karami, "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*, vol. 43, no. 1, pp. 53-67, 2012.
- [6] I. Elgort, "Effects of L1 definitions and cognate status of test items on the vocabulary size test," *Language Testing*, vol. 30, no. 2, pp. 253-272, 2013.
- [7] J. M. Linacre. (2010). A User's Guide to WINSTEPS. [Online]. Available: <http://Chicago:winsteps.com>



**R. Derrah** is originally from Boston, Massachusetts, U.S.A. He graduated with a bachelor's degree in history from Bridgewater State University in 1996. He also holds a master's degree in education from Boston College and a Master's degree in east Asian studies from Harvard University. He is currently a PhD candidate at Temple University.

He has taught both in the United States and Japan, and currently holds a position at Kwansai Gakuin University in Hyogo Prefecture, Japan. His current research interests include secondary education in east Asia and well as secondary level teacher training and certification.

Prof. Derrah is a member of the Association of Asian Studies.



**D. Edward Rowe** is originally from Cincinnati, Ohio, U.S.A. He graduated with a bachelor's degree in mass communications from Miami University, Ohio in 1994. He is currently working on his master's degree in education at Temple University, Osaka, Japan and resides in Japan.

He has taught TOEIC and business communication for many companies in Japan including: Mitsubishi Heavy Industries, Takeda Pharmaceuticals, and Sumitomo Chemical. He is currently employed by Kobe Women's University and Keimeji High School, both in Kobe. He recently published an article titled "Accommodating different perspective learning styles" in *Individual Differences in Second Language Acquisition*, vol. 94, 34-39, August 2014.