# Study on the Validity of bilingual Mandarin Version of Vocabulary Size Test

Yu Wang[1] & Wanyi Du[1]

[1] School of Software, Dalian University of Technology, Liaoning, China

Correspondence: Yu Wang, School of Software, Dalian University of Technology, Dalian Economy & Technology Development Zone, Dalian 116620, China. Tel: 86-159-0425-9125. E-mail: karan_wang@dlut.edu.cn

## Abstract

Vocabulary Size Test (VST) is a test to measure learners' receptive vocabulary size by Paul Nation. Beside English-English version, VST has other five bilingual versions. This article is to describe the validation of its Mandarin version, i.e., English-Chinese version. Data shows that the bilingual English-Chinese version VST's overall difficulty index, discriminating power, and internal consistency are all statistically satisfactory. One-way ANOVA test and Post Hoc test evidenced that the English-Chinese version VST can differentiate learners from different levels with high validity.

**Keywords**: English-Chinese bilingual version VST, vocabulary size test, validity

## 1. Introduction

Meara (1996) pointed out "learner with large vocabulary is more proficient in each language skills than learners with small vocabulary, which suggests that vocabulary is an important factor in language."Studies on vocabulary size attracted increasing attention from researchers in China. Unfortunately, problem arose that the test results differ noticeably with each other (Zhou et al., 2008). Shao (2002) found that the vocabulary size of freshman in a normal university was 2547 and it increased to 3811 at the end of second year. Lu (2004) investigated that the freshmen in a leading science and engineer university in southwestern China had a vocabulary size of 2145. Cui and Wang (2006) found that the receptive vocabulary size of the undergraduate student majored in English was 3391 in their first year and increased to 7199 in the fourth year. Lu (2004) pointed out that the different vocabulary testing tool accounted for the inconsistence in test results. Zhou et al (2008) argued that several factor had to be considered to ensure the validity of a scientific testing research, namely, the test designers' definition of vocabulary size and vocabulary ability, sampling approach and ample resource etc. Tang and Han (2010) posited that, currently in China, researchers adopted different vocabulary testing tools without a universally-recognizable standard, which lead to immaturity in the application of testing tool. Obviously, the study about the validity of different vocabulary testing tools is not adequately emphasized.

## 2. Literature Review

In order to test the vocabulary size of EFL learner, Meara and Buxton (1987) developed the Yes/No Vocabulary Checklist. However, it is not ideally applicable to all second language learners. For example, the Arabic or French learner studying English as their second language produced a higher false alarming rate in the test (Cobb, 2000). Later, Nation (1983, 1990) designed Vocabulary Level Test (hereinafter called VLT) for Asian second language learners. Nation, the designer of VLT, also admitted that the function of VLT is to diagnose rather than to test the vocabulary size of a ESL learner.

In 2007, Nation and Beglar teamed up to develop Vocabulary Size Test, VST. It includes 14 vocabulary frequency levels, composing of 10 sets of questions each and totaling 140 questions. VST include an English-English monolingual version and five bilingual versions, namely, bilingual Korean version, bilingual Japanese version, bilingual Mandarin version, bilingual Russian version and bilingual Vietnamese version (all can be traced on http://www.victoria.ac.nz/lals/about/staff/paul-nation). Compared with VLT in which the answer can be easily guessed, VST is more difficult and controllable in the multiple choice questions, which ensure the maximal validity and reliability (Nation & Beglar, 2007). The monolingual English-English VST proved to be

statistically highly valid in differentiating various levels of English learners as exhibited by Rasch model-based data analysis (Beglar, 2010).

This paper is to study the validity of bilingual Mandarin version of VST, i.e., whether English-Chinese bilingual VST (herein referred to as E-C VST) is as effective as the bilingual English-English VST (herein referred to as E-E VST) in testing the receptive vocabulary of Chinese students. The followings are the research questions:

1)  Whether the degree of difficulty, differentiation and reliability of E-C version of VST meet the statistics standard.

2)  Whether E-C version of VST could effectively differentiate language levels of second language learners? In other words, whether the receptive vocabulary of the learners with different levels differs significantly?

3)  What is the criterion related to validity between E-C VST and English-English VST?

## 3. Method

### 3.1 Participants

90 sophomore students, 62 boys and 28 girls participated in this research. They are all software–majored students with same academic background and similar English-learning experience.

### 3.2 Grouping Procedures

Participants were grouped according to their scores in national College English Test (CET).

CET is recognized as the most valid national English test in China, held twice a year across the country. CET includes two levels: the CE4 and the advanced CET6. Certificate of CET4 was once regarded as an indispensable condition to earn a bachelor degree in China. In 2008, the structure of CET4 was dramatically reformed to meet international testing standard and CET4 was stopped to be bundled together with bachelor diploma since then. Full score of CET4 is 710, 550 for excellence and 425 as pass line.

Most current studies on vocabulary test group subjects based on tests rather than new version of CET4. The application of new CET4 test as placement tool is believed to be the innovation of this study.

We group all 406 sophomore students according to their CET scores, namely, 580-710 as high level group, 490-520 as intermediate level group and 350-425 as low level group. Then, we randomly picked 30 from each group as subjects in this study.

### 3.3 Test Procedure

90 participants from three groups were first required to finish 14 levels of E-C VST within two hours. The test was monitored by researcher to ensure a no-dictionary, no-reference material and no discussion testing environments. The E-E VST was conducted two weeks later under same condition. The test papers were graded by researcher manually and grades were entered into SPSS for further analysis.

## 4. Results and Discussions

### 4.1 Whether the Degree of Difficulty, Differentiation and Reliability of E-C Version of VST Meet the Statistics Standard

Table 1. The difficulty, differentiation and reliability of E-C VST

| Order | Difficulty | Differentiation | Difficulty Degree | Difficulty | Differentiation |
|---|---|---|---|---|---|
| 1st 1000 vocab level | 0.89 | 0.25 | 8th 1000 vocab level | 0.38 | 0.32 |
| 2nd 1000 vocab level | 0.72 | 0.28 | 9th 1000 vocab level | 0.3 | 0.44 |
| 3rd 1000 vocab level | 0.55 | 0.40 | 10th 1000 vocab level | 0.36 | 0.36 |
| 4th 1000 vocab level | 0.5 | 0.36 | 11th 1000 vocab level | 0.42 | 0.43 |
| 5th 1000 vocab level | 0.45 | 0.46 | 12th 1000 vocab level | 0.35 | 0.32 |
| 6th 1000 vocab level | 0.46 | 0.36 | 13th 1000 vocab level | 0.29 | 0.40 |
| 7th 1000 vocab level | 0.47 | 0.40 | 14th 1000 vocab level | 0.22 | 0.46 |
| Internal Consistency Reliability (Alpha): 0.905 | | | | | |

Table 1 lists the Difficulty Degree, Differentiation and reliability of 14 vocabulary levels. Data indicates that all of the three indicators of E-C VST meet the statistics standard. The difficulty degree of $1^{st}$ 1000 vocabulary level in this testing tool is 0.89, and gradually decreases with the increase of vocabulary difficulty. The difficulty degree of $7^{th}$ 1000 vocabulary is 0.47, the $14^{th}$ 1000 vocabulary is 0.22. This testing tool intends to test the vocabulary quantity of each vocabulary of each vocabulary level that students have acquired. Test begins from the vocabulary of high frequency to the vocabulary of low frequency with an increase of vocabulary difficulty. The data shows that the gradient is reasonable and conforms to the statistics standard from the respect of difficulty degree. Psychometrist Eble pointed out that the high differentiation of test question should be above 0.4, and 0.2-0.29 indicates low differentiation. Table1 shows that the differentiation of the first two vocabulary levels in this testing tool are 0.25 and 0.28, respectively, a relatively low differentiation because these two vocabulary levels are the most basic and frequent English vocabulary that students must master. Starting from $3^{rd}$ vocabulary level, the differentiation fluctuates between 0.32-0.46 with the increase of vocabulary difficulty.

*4.2 Whether E-C VST Could Effectively Differentiate Second Languages Learners' Level*

Table 2 lists the test score of learners of low, middle and high level in the test of E-C version of VST, which could be regarded equally as the receptive vocabulary size of Chinese students. We grouped the students according to their score in the new version of CET4.CET4 is currently the most widely influential language test, which means that both its validity and reliability are supposed to be mostly close to the standard of large-scale standardized test.

Table 2. Descriptive statistics of the test score of students of low, middle, high level students in E-C VST

| | CET score | subjects | Mean | SD | Max. | Min. |
|---|---|---|---|---|---|---|
| High level | 580~660 | 30 | 7393.33 | 1320.249 | 10850 | 4950 |
| Middle level | 490~510 | 30 | 6523.33 | 1272.842 | 11500 | 5300 |
| Low level | 350~425 | 30 | 5178.33 | 920.708 | 6400 | 3150 |
| Total | | 90 | 6385.00 | 1487.163 | 11500 | 3150 |

Table 2 shows that according to E-C VST, students' receptive vocabulary of different English level are: high level group/7393; middle level group/6523; low level group/5178. As mean score presents, this test tool could substantially differentiate the vocabulary size of students from different level. The standard deviation of high level and middle level students are higher than that of low level students. Comparatively, there is less individual difference among low level students. Accordingly to the post-hoc analysis of the three groups, the vocabulary of middle level students is 25% higher than that of low level students; the vocabulary of high level students is 13% higher than that of middle level students. These data adequately proves that E-C VST could effectively differentiate the receptive vocabulary size of different levels students.

In the College English Curriculum Requirement issued in 2007, the general requirement of recommended vocabulary in college English curriculum in China is 4795 words for basic level, 6395 words for intermediate level and 7675 words for higher level. The CET4 score of the three groups of students in this study generally represents the different English level of college students in China.

Table 3. ANOVA of test score of three groups of students in E-C VST

| Variance Resource | Sum of Square of Deviation | Df. | Mean square | F-value | P-value |
|---|---|---|---|---|---|
| 14000 total words | | | | | |
| BS | 74721500 | 2 | 37360750.00 | 26.617 | .000 |
| WS | 1E+008 | 87 | 1403629.310 | | |
| Total | 2E+008 | 89 | | | |
| First 7000 words | | | | | |
| BS | 26805722 | 2 | 13402861.11 | 31.437 | .000 |
| WS | 37092083 | 87 | 426345.785 | | |
| Total | 63897806 | 89 | | | |
| 7000-14000 words | | | | | |
| BS | 12040389 | 2 | 6020194.444 | 10.920 | .000 |
| WS | 47965167 | 87 | 551323.755 | | |
| Total | 60005556 | 89 | | | |

In order to study the significant difference among three groups, we conducted ANOVA and Scheffe post-hoc analysis. Table3 shows that three groups performance differ significantly and the result of Scheffe Post-hoc also shows significant difference in-between each group. To further study the significance between high and low frequency vocabulary among three groups of students, we conducts ANOVA and post-hoc on the first 7000 words (high frequency words) and 7000-14000 words (low frequency words). Results exhibits that the test score of three groups differ significantly for the high frequency words, post-hoc also shows significant difference in-between each group (P<.05). However, although there is statistically-significant difference for the three groups' test score on low frequency words, post-hoc show no significant difference between each group (P=.25). It means that E-C VST could not reflect the difference among students of different levels with middle and low frequency words. In China, according to College English Curriculum Requirement, the requirement for high level is 7675 words for College students. For most of the test takers, they won't be familiar with the vocabulary at 7000-14000 level, which also explains the insignificant difference between the middle and high level students with the test of low frequency vocabulary.

*4.3 What Is the Criterion Related Validity between E-C Version of VST and E-EE-E Version of VST*

We see E-E VST as criterion and test the simultaneous validity of E-C VST by comparing test score of three groups of both E-E VST and E-C VST. Students took E-E VST two weeks after the completion of E-C VST. Both tests are completed in classroom and the requirements of them are consistent in order to exclude the negative influence of the disregard and inadequate seriousness of the students. We conducted Pearson correlation analysis about the results of the two tests.

Table 4. Pearson correlation analysis of the vocabulary E-E VST score of three groups of students

| | | High level group (n=30) | | Middle level group (n=30) | | Low level group (n=30) | |
|---|---|---|---|---|---|---|---|
| | | E-E VST Correlation | P value | E-E VST Correlation | P value | E-EVST Correlation | P value |
| E-C VST | 1000-7000 level | .652 | .000 | .822 | .000 | .270 | .149 |
| | 7000-14000level | .638 | .000 | .488 | .006 | .405 | .026 |
| | Total | .711 | .000 | .711 | .000 | .287 | .124 |

Table 4 shows that for high level students, there is a significant positive correlation between the test scores of high frequency and low frequency words in both E-C and E-E VST (.652,.638, p=.000). For middle level students, there is significant positive correlation between the test scores of low frequency vocabulary of E-C VST and that of E-E VST (.822, p=.000). There is statistically significant correlation in the test of high frequency vocabulary (.488, p=.006), which is relatively low compared with that of low frequency vocabulary. For low level students, there are no significant correlations. Previous study proved that it is an effective method to use first language to test the vocabulary of second languages for low level students (Nation, 2001). While on the other hand, according to psycholinguistic study, high-level bilingual students could grasp their second languages as well as their first languages so that two languages share resources sometimes (Magiste, 1984, cited from Lu & Tu, 2010).

**5. Conclusion**

Vocabulary ability is essential in language skill. Vocabulary test not only differentiates students' vocabulary ability, but also provides data in recognizing ESL learners' language ability. In 2007, Nation and Beglar worked out the Vocabulary Size Test (VST), containing14 vocabulary frequency levels and a total of 140 questions. This paper analyzed the validity of the bilingual Mandarin version of VST, namely, the English-Chinese bilingual VST (E-C VST).

We adopted new version of CET4 as its grouping standard and conducted statistical analysis on difficulty degree, ANOVA and post-hoc to test the validity of E-C VST. Results show that the degree of difficulty of E-C VST demonstrated gradual decreases in a reasonable trapezoid shape. The results of ANOVA indicate good validity of E-C VST, which could differentiate the receptive vocabulary of students of different levels. However, the results of post-hoc show that for middle and high-level students, there is no significant difference in the recognition rate of low frequency vocabulary. The result of simultaneous validity of E-E VST and E-C VST indicates that E-C VST is more applicable for low level students. Whereas, high level students demonstrate consistent performance in both E-E VST and E-C VST, which means that E-E VST and E-C VST bore simultaneous validity both to high-level learners.

This research is preliminary in studying the validity of E-C VST. Considering the number of test takers, large–scale follow-up study is necessary in further study. Further study would group different levels of learners according to listening, speaking, reading and writing skills respectively so as to discover the correlation between specific language skills and vocabulary size.

**Acknowledgment**

**Reference**

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101-18. http://dx.doi.org/10.1177/0265532209340194

Cobb, T. (2000). One size fits all? Francophone learners and English vocabulary tests. *Canadian Modern Language Review, 57*(2), 295-324. http://dx.doi.org/10.3138/cmlr.57.2.295

Cui, Y., & Wang, T. (2006). Study on the development route and correlation of receptive vocabulary, productive vocabulary and depth of vocabulary knowledge. *Modern Foreign Language, 29*(4), 392-400.

Lu, Z. H., & Tu, L. (2010). The Age of second language acquisition and mental representation of highly proficient C-E bilinguals. *Foreign Languages, 33*(4), 47-56.

Lv, C. H. (2004). The relationship between vocabulary and comprehensive language skill, the depth of vocabulary knowledge. *Foreign Language Teaching and Researching, 36*(2), 116-123.

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

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524759

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Shao, H. (2002). Empirical study on the CET-4 vocabulary level of the students in normal university. *Foreign Language Teaching and Research, 34*(6), 421-425.

Tang, W., & Han, B. (2010). The Application of vocabulary testing tool in the domestic correlation study. *Foreign languages Teaching and Researching, 18*(2), 87-91.

Zhou, S., Gu, Y., & Guo, Y. (2008). Review of Vocabulary Test. *Foreign Language World, 2008*(1), 90-95.

**Copyrights**