

Lexical Testing and the Reliability of the Modified Vocabulary Knowledge Scale

Debbita Tan (Corresponding author)

School of Languages, Literacies and Translation, Universiti Sains Malaysia, Malaysia

E-mail: debbita_tan@usm.my

Ambigapathy Pandian

School of Languages, Literacies and Translation, Universiti Sains Malaysia, Malaysia

Paramaswari Jaganathan

School of Languages, Literacies and Translation, Universiti Sains Malaysia, Malaysia

Doi:10.7575/aiac.all.v.7n.5p.91

Received: 28/05/2016

URL: <http://dx.doi.org/10.7575/aiac.all.v.7n.5p.91>

Accepted: 01/08/2016

Abstract

This paper highlights some of the vocabulary tests available, and reports the reliability of the modified Vocabulary Knowledge Scale (VKS) (Rosszell, 2007). Although there is no consensus as to what actually constitutes vocabulary knowledge, the notion that it is made up of receptive knowledge (words recognised or known when seen or heard) and productive knowledge (words appropriately used when we write or speak) is widely accepted. Lexical testing is important for various reasons, chiefly to determine reading ability which requires the use of a size test, and to monitor overall vocabulary development which necessitates the use of a test measuring both receptive and productive knowledge (for instance, the modified VKS). The modified VKS was pilot-tested on 28 university-level Malaysian remedial English language learners and analysed for 'reliability as stability over similar samples'. Data analyses returned values indicating the test to be reliable, thus presenting it as a feasible option for use among similar cohorts. This is of significance to scholars, researchers, language instructors and curriculum designers intending to employ the test in their own research, classrooms and literacy programmes.

Keywords: lexical testing, receptive and productive vocabulary knowledge, modified Vocabulary Knowledge Scale, reliability, tertiary learners

1. Introduction

1.1 Vocabulary Knowledge

Vocabulary knowledge has been defined differently by different researchers, with varying opinions as to what it means to *know* a word. Anderson and Freebody (1981) parodied this by stating the following:

"It is not clear that, if Ludwig Wittgenstein and Bertrand Russell were left alone in a room for three hours, they could decide that they really knew the meaning of 'dog'." (p. 90)

Schmitt and Meara (1997) grouped vocabulary knowledge into three broad dimensions: form, meaning and use. Henriksen (1999), meanwhile, in tandem with the position that vocabulary knowledge is multidimensional, put forth that vocabulary knowledge comprises three components: 1) partial-to-precise knowledge (varying degrees of understanding), 2) depth-of-knowledge (demonstrating word knowledge's multifaceted nature), and 3) receptive-productive dimension (an individual's comprehension and production abilities).

Fundamentally, vocabulary knowledge comprises two forms: the receptive dimension and the productive dimension. The former consists of words that we know when we see or hear them, whereas words that we use appropriately when we write or speak are considered productive vocabulary knowledge (Lehr, Osborn, & Hiebert, 2004).

With regards to receptive vocabulary knowledge, Nation (1990) suggested that its more commonly accepted feature is the ability to recognise word form and retrieve meaning in reading and listening. As for the productive dimension, it is the ability to use words correctly to express a certain meaning through writing or speaking (Nation, *ibid.*). Additionally, Nation (2001) remarked that a learner's understanding of a particular word is reflected in his or her usage of it.

1.2 Lexical Testing

Laufer and Goldstein (2004) pointed out that very few vocabulary tests attempt to measure a learner's progress along a continuum of knowledge. An exception is Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS), which

tracks a learner's progress from total unfamiliarity to superficial familiarity with a lexical item to the ability to use it accurately in production.

Belisle (2000) noted the researchers' reference to Cronbach's categories of increasing knowledge of words (developed in 1942):

1. *Generalisation* – able to define the word
2. *Application* – selecting an appropriate use of the word
3. *Breadth of meaning* – recalling the different meanings of the word
4. *Precision of meaning* – applying the word correctly to all possible situations
5. *Availability* – able to use the word productively

Paribakht and Wesche (1997) described the Vocabulary Knowledge Scale as a practical instrument that can be used with any set of words, and useful for studies concerned with the recognition and use of words. The VKS differs from other scale tests mainly in terms of *verifiable* evidence of knowledge; there are provisions in the VKS which prompt testees to *demonstrate* their receptive knowledge of a particular word as well as their ability to use it (productive knowledge). The format of the VKS is as follows:

1	I don't remember having seen this word before.
2	I have seen this word before, but I don't think I know what it means.
3	I have seen this word before, and I think it means _____ [synonym or translation]
4	I know this word. It means _____ [synonym or translation]
5	I can use this word in a sentence: _____

Figure 1. Vocabulary Knowledge Scale Format (Wesche & Paribakht, 1996)

Over the years, the VKS has gained considerable currency in ESL/EFL vocabulary knowledge assessment and the scale test as well as its variants have been used in numerous studies (Waring, 2002). In order to better capture partial knowledge of word meaning, Rosszell (2007) added the following statement in the VKS: 'I have seen this word before and I think it is related to the following word/idea: _____'. In addition, to facilitate better evaluation of the English sentences produced by testees, the requirement that each sentence be translated into the testees' L1 (in the case of Rosszell's study, Japanese) was added (see Appendix).

This modified version of the Vocabulary Knowledge Scale was adapted and utilised in the present research and tested for reliability. The following is a sample item:

Word: anxious	
Form: Adjective	
1]	I do not think I have ever seen this word. Tick (✓) if true: _____ (do not proceed)
2]	I have seen this word before, but I do not know what it means. Tick (✓) if true: _____ (do not proceed)
3]	I have seen this word before and I think it is related to the following word/idea: _____ (answer may be given in English/Bahasa Malaysia)
4]	I have seen this word before and I think it means: _____ (give a synonym or definition in English/Bahasa Malaysia)
5]	I cannot use this word (anxious) in a sentence. Tick (✓) if true: _____ (do not proceed)
6]	I can use this word (anxious) in a sentence: _____ (write your sentence in English)
7]	Translate your sentence into Bahasa Malaysia: _____

Figure 2. Modified Vocabulary Knowledge Scale (Adapted Version – Sample Item)

Scoring Guide

A response at Level 1 yields a score of 0 while a response at Level 2 yields a score of 1. A related word or idea provided at Level 3 yields a score of 2, but an unrelated or incorrect one yields a score of 1. If a correct synonym or definition is provided at Level 4, a score of 3 is awarded (an incorrect answer yields a score of 1). A response at Level 5 yields a score of 0. If no English sentence is provided at Level 6, a score of 0 is awarded (the same score is awarded if the sentence is incomprehensible or if the use of the target word is semantically inappropriate). However, if the translated sentence at Level 7 clearly shows an understanding of the meaning of the target word, a score of 1 is awarded. If the sentence at Level 6 demonstrates knowledge of the target word but is grammatically incorrect or is a different part of speech, a score of 2 is awarded. If the sentence at Level 6 is both grammatically and semantically correct, a score of 3 is awarded.

It has been suggested that scale tests present a better analysis than what can be achieved with a standard yes/no test (Waring, 2002), a position in tandem with Read's (2000) following statement:

"all that we can confidently say about a "yes" response to a word is that the learner is familiar with the word form and can identify it as a real English word" (p. 148)

Examples of vocabulary tests that measure just one form of lexical knowledge are Nation and Beglar's (2007) Vocabulary Size Test, which measures only receptive knowledge and does so through word association, and Meara and Buxton's (1987) Yes/No Vocabulary Test which requires testees to indicate which of the words in a list are known to them. Both tests enjoy vast recognition and application (Eyckmans, 2004), being convenient to administer and practical for use in research involving large samples. Such tests are also extremely useful for admission purposes and more accurate placements in language learning courses (e.g., reading programmes), as well as for the purpose of monitoring a learner's growth in terms of receptive knowledge.

Another example is Read's (1998) Word Associates Test (WAT). The WAT emphasises the form-meaning link and presents testees with a stimulus accompanied by four possible synonyms and four possible collocates, from which four correct associates should be selected:

1. <i>beautiful</i>							
enjoyable	expensive	free	loud	education	face	music	weather

Figure 3. Word Associates Test (Sample Item) (Read, 1998)

On the whole, despite research describing vocabulary knowledge as multidimensional, many tests continue to stress on the form-meaning link. Admittedly, this is inevitable as words are primarily units that represent meanings and the very objective of vocabulary learning is, foremost, to develop a mental database of meaningful words from which one can retrieve lexical information for use in activities such as reading in the target language. However, vocabulary tests in which the form-meaning link is made central without testing for production remain essentially assessments that solely determine receptive vocabulary knowledge; the absence of the productive component renders the results of such tests unrepresentative of their testees' *overall* lexical ability.

In essence, vocabulary tests are designed and prescribed according to different conditions, these primarily being the test designer's views on and definition of vocabulary knowledge, and the purpose of administering a vocabulary test in the first place. If the objective is to determine learners' vocabulary size before the start of a reading programme, then tests such as Nation and Beglar's (2007) Vocabulary Size Test would suffice. On the other hand, if the aim is to investigate lexical ability in a more comprehensive manner or to have a more representative picture of the effects of a vocabulary learning intervention, then two-dimensional tests like the VKS (Roszell, 2007; Wesche & Paribakht, 1996) would be a more consonant choice.

1.3 Testing for Reliability

1.3.1 Previous Findings

Wesche and Paribakht's (1996) study of the VKS reported high correlations (above .90) between learners' self-report of word knowledge and actual score for demonstrated knowledge. Additionally, their administration of a test-retest format garnered readings of above .80, denoting reliability. The study was set within a university ESL context involving learners of different proficiency levels.

In his research involving tertiary-level Japanese EFL students, Roszell (2007) found generally high internal reliability coefficients for each administration of the modified VKS. He also reported interrater reliability rates of more than 90% agreement, indicating reliability of the scoring/marking scheme.

Despite these encouraging findings, however, it is pertinent to bear in mind that although reliability can be estimated through the research of others, reliability assessment remains important when language testing instruments are to be administered on new types of subjects or cohorts.

1.3.2 Current Analysis

According to Cohen, Manion and Morrison (2007), in their comprehensive work on research methods in education, reliability in quantitative research is basically a synonym for these: dependability, consistency, replicability. Reliability is a measure of stability (e.g., over similar samples) (Cohen, Manion, & Morrison, *ibid.*) and as such, although reliability can be estimated through existing research, reliability assessment remains essential when language testing instruments are to be administered on new types of cohorts.

The present paper reports on the reliability of the modified Vocabulary Knowledge Scale (Rosszell, 2007). To the researchers' knowledge, there is no published research on the reliability of this test specifically within the context of its use among university-level Malaysian remedial English language learners grouped under the lower proficiency MUET (Malaysian University English Test) bands of 1 to 3.

The MUET is a prerequisite for admission into Malaysian public universities and is managed by the Malaysian Examinations Council; it has been validated as a reliable measure of English proficiency (the lowest proficiency band is '1' and the highest is '6') (Souba & Kee, 2011).

AGGREGATED SCORE	BAND	USER
260 – 300	6	Highly proficient
220 – 259	5	Proficient
180 – 219	4	Satisfactory
140 – 179	3	Modest
100 – 139	2	Limited
Below 100	1	Very limited

Figure 4. MUET Band Description

2. Methodology

2.1 Participants

The modified Vocabulary Knowledge Scale was pilot-tested and analysed for 'reliability as stability over similar samples'. Cohen, Manion and Morrison (2007) put forth that dependability, consistency and replicability fundamentally represent reliability; a reliable instrument should yield similar results when similar participants are involved. It is assumed that should a reliable test be administered on different groups of individuals who are comparable on factors that have a significant bearing on the test results, then similar results would be acquired; reliability can be determined using the t-test, with a significance level of .05 or higher indicating reliability (Cohen, Manion, & Morrison, *ibid.*).

Two groups (A and B) of tertiary-level Malaysian remedial English language learners were involved, with each group made up of 14 participants who were registered for a preparatory English proficiency course at a public university in Malaysia. They were between the ages of 19 to 22 years old and grouped under the lower proficiency MUET bands of 1 to 3.

2.2 Procedure

Consent was sought and obtained, and the test was administered on both groups at the beginning of the academic semester. Before the start of the test, the participants were briefed on how to complete it and a two-hour time limit was set. A total of 30 words were tested, comprising nouns, adjectives, verbs and adverbs.

3. Results

The test measures both receptive and productive vocabulary knowledge. Section (i) presents results pertaining to the receptive component and Section (ii), the productive component.

Section (i)

Table 1. Results: Receptive vocabulary knowledge

Group	<i>M</i>	<i>SD</i>
A	65.21	8.01
B	65.86	8.40

Table 1 shows the mean scores for both groups; A ($M=65.21$, $SD=8.01$) and B ($M=65.86$, $SD=8.4$). Baseline similarity is indicated.

Table 2. T-test output: Receptive vocabulary knowledge

	Sig.
Receptive vocabulary knowledge: A-B	.837

Table 2 shows the Sig. (p) value obtained to be higher than .05 ($p>.05$) at $p=.837$, indicating no statistically significant differences between group means, thus denoting reliability in terms of stability over similar samples.

Section (ii)

Table 3. Results: Productive vocabulary knowledge

Group	<i>M</i>	<i>SD</i>
A	19.79	2.89
B	20.43	2.56

Table 3 shows the mean scores for both groups; A ($M=19.79$, $SD=2.89$) and B ($M=20.43$, $SD=2.56$), indicating baseline similarity.

Table 4. T-test output: Productive vocabulary knowledge

	Sig.
Productive vocabulary knowledge: A-B	.539

Table 4 shows the Sig. (p) value obtained to be higher than .05 ($p>.05$) at $p=.539$, indicating no statistically significant differences between group means, thus denoting reliability in terms of stability over similar samples.

On the whole, the results indicate that the test is a reliable instrument for use with similar cohorts (i.e., university-level Malaysian remedial English language learners grouped under the lower proficiency MUET bands of 1 to 3, or equivalent).

4. Discussion and Conclusion

The findings have demonstrated the modified Vocabulary Knowledge Scale (Rosszell, 2007) to be a reliable testing instrument for use within the ESL/EFL context. As the form of reliability reported here is within the framework of stability over similar samples, the results are therefore not extended to include cohorts that are external of the defined parameters. The findings, however, are most relevant specifically to cohorts as defined in the present paper as well as to cohorts that fall within equivalent parameters.

The present findings are of significance to scholars, researchers, language teachers and curriculum designers intending to employ the test in their own research, classrooms and literacy programmes. The p -values obtained from the t-tests are highly encouraging at .837 and .539 – these values far exceed the pre-determined significance level of .05, an indication of good reliability.

In sum, the modified VKS represents an option that is reliable for use among university-level Malaysian remedial English language learners grouped under the lower proficiency MUET bands of 1 to 3, or equivalent. Additionally, it is to be noted that apart from its reliability, the test also represents a feasible option due to its usability (ease of use) in terms of test delivery.

References

- Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.
- Belisle, T.A. (2000). Developing vocabulary knowledge in the immersion classroom. *The bridge: From research to practice*. Edina, MN: Normandale French Immersion School.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education (Sixth Edition)*. New York, NY: Routledge.
- Eyckmans, J. (2004). *Measuring receptive vocabulary size*. Utrecht, NL: LOT.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Lehr, F., Osborn, J., & Hiebert, E.H. (2004). *A focus on vocabulary*. Honolulu, HI: Pacific Regional Educational Laboratory.
- Meara, P., & Buxton, P. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142-154.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York, NY: Newbury House.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Paribakht, T.S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 225-237). Cambridge, UK: Cambridge University Press.
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41-60). Mahwah, NJ: Lawrence Erlbaum.
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- Roszell, H.R. (2007). *Extensive reading and intensive vocabulary study in a Japanese university*. (Unpublished doctoral thesis). Temple University, Japan.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word association and verbal suffixes. *Studies in Second Language Acquisition*, 20, 17-36.
- Souba Rethinasamy, & Kee, M.C. (2011). The Malaysian University English Test (MUET) and its use for placement purposes: A predictive validity study. *Electronic Journal of Foreign Language Teaching*, 8(2), 234-245.
- Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. Appeared in *Kiyo: The Occasional Papers of Notre Dame Seishin University*, Japan.
- Wesche, M., & Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40.

APPENDIX

Example of Modified VKS Item (Roszell, 2007)

warrant (n.)	
i)	I don't think I have ever seen this word. _____
ii)	I have seen this word before, but I don't know what it means. _____ (For levels iii) and iv), write in English or Japanese hiragana.)
iii)	I have seen this word before and think that it is related to the following word or idea: _____
iv)	I have seen this word before and I think it means _____
v(a)	I can use this word in a sentence. (Write a sentence.) _____ _____
(b)	(If you write a sentence, you must also write the meaning in iii) and/or iv). Your English sentence translated into Japanese (in hiragana). _____ _____