Assessing Reliability of Two Versions of Vocabulary Levels Tests in Iranian Context

Aso Bayazidi

Department of English, Boukan Branch, Islamic Azad University, Boukan, Iran

E-mail: bayazidi.aso@gmail.com

Fateme Saeb (Corresponding author)

Faculty of Foreign Languages, University of Isfahan, Isfahan, Iran

E-mail: F.saeb1384@gmail.com

Abstract

This study examined the equivalence and reliability of the two versions of the Vocabulary Levels Test in an Iranian context. This study was motivated by the fact that the Vocabulary Levels test is increasingly being used in Iran for both research and pedagogical purposes without having been checked for validity and reliability in this context. The equivalence and reliability of the two versions of the test were examined through the parallel-form approach to reliability in Classical True Score theory. Seventy-five intermediate learners of English as a foreign language at the Iran Language Institute took the two versions of the test with one week interval between the two administrations in a counterbalanced fashion. To examine the equivalence of the two versions, the means and variances of the scores obtained for the two tests were compared using paired-sample t-test and one-way ANOVA, respectively. The results of the analyses indicated that the difference between the means of the two versions was significant, and the two versions cannot be considered as parallel forms. To assess the reliability of the two versions, the correlation between the scores obtained from them was estimated using Pearson Product Moment correlation. The results of the analyses showed that the two versions are highly correlated and are reliable tests. It is concluded that the two versions should not be treated as equivalent in longitudinal and gain score studies.

Keywords: Vocabulary Testing, The Vocabulary Levels Test, Equivalence, Reliability

1. Introduction

Vocabulary is an important component of language use. As Wilkins (1972) states, "without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p. 111). Lack of vocabulary knowledge impedes English as a Second Language (ESL) learners' ability to comprehend messages or express themselves clearly in English. There is plenty of evidence pointing to the importance of vocabulary in language use.

One of the most systematic explorations of the relationship between vocabulary knowledge and language proficiency occurred as part of the development of the DIALANG tests (Alderson, 2005, as cited in Schmitt 2010). His research team compared scores on various vocabulary tests with the scores from the other language components of the DIALANG test. The results revealed that vocabulary has strong relationships with the language skills. The checklist test and the vocabulary test battery correlate with reading at .64, listening from .61-.65, writing from .70-.79, and grammar at .64. Several other studies have highlighted the importance of vocabulary knowledge for second language (L2) learners in reading (Haynes & Baker, 1993; Huckin & Bloch, 1993), speaking (Joe, 1998), listening (Elley, 1989; Ellis, 1994), and writing (Lee, 2003; Hinkel, 2001, Laufer & Nation 1995). What the above studies would appear to show is that vocabulary knowledge contributes to a great deal to overall language success.

Given the importance of vocabulary in language use, one key issue in vocabulary studies is how much vocabulary is necessary to enable communication. The short answer is a lot, but it depends on one's learning goals. If one wishes to achieve native-like proficiency, then presumably it is necessary to have a vocabulary size similar to native speakers. There have been a few well-designed studies which provide reliable estimates of native language vocabulary size. Goulden, Nation, and Read (1990) found that their New Zealand university undergraduates had a vocabulary size of about 17,000 word families. D'Anna, Zechmeister, and Hall (1991) found that their university students knew a little under 17,000 of the headwords in the 1980 *Oxford American Dictionary*. Schmitt (2010) remarks that despite the fact that native speakers will always vary in their vocabulary size to some extent, a range of 16,000-20,000 word families seems a fair estimate of the vocabulary size for educated native speakers.

Luckily, L2 learners do not need to achieve native-like vocabulary sizes in order to use English well. A more reasonable vocabulary goal for these learners is the amount of lexis necessary to enable the various forms of communication in

English. One of the most basic things a person might want to do is to communicate orally on an everyday basis (e.g. asking directions to the train station, describing one's holiday). If we assume that 98% of the vocabulary needs to be known (Hu and Nation, 2000), we can estimate the number of word families it takes to be able to engage in informal daily conversation. According to Nation (2006) a base of 6,000-7,000 word families is needed to meet this goal. He also estimated that 95% coverage would require knowledge of about 3,000 word families, plus proper nouns. Overall, the current evidence suggests that it requires between 2,000 and 3,000 word families to be conversant in English (if 95% coverage is adequate) or between 6,000 and 7,000 word families if 98% coverage is adequate. However, there is not enough evidence to confidently establish a coverage requirement for listening at the moment. We are on firmer ground for estimates of written vocabulary. To further complete the picture of L2 vocabulary size, Nation (2006) used The British National Corpus (BNC) data and 98% coverage to calculate that 8,000-9,000 word families are required to read authentic texts (e.g. novels or newspapers) in English. As such, it makes sense to be able to measure learners' knowledge of vocabulary.

The need for reliable and valid tests of vocabulary size is a critically important issue in the field of second language acquisition (SLA). This is equally true whether we are interested in pedagogical assessment in classrooms or in language acquisition research. Given this, one might expect there to be an accepted vocabulary test available for these uses. However, this is not the case yet. In lamenting the lack of reliable tests for measuring vocabulary size, Meara (1996) states that the nearest thing the field has to a standard test in vocabulary is the Vocabulary Levels Test (Nation, 1983, 1990). Different versions of the test have been widely employed in both assessment and research around the word, and it is staring to be used for different purposes in Iran. Despite this widespread use, the test has been properly checked for reliability and validity by very few studies. This article aims at examining the reliability of the test in an Iranian context.

2. Review of the Literature

2.1 Reliability

A language test as a measuring instrument is required to generate individual scores that are reliable and valid. Ary, Jacobs, and Sorensen (2010) define reliability of a measuring instrument as the "degree of consistency with which it measures whatever it is measuring" (p. 236). Reliability means that scores from an instrument are stable and consistent. Scores should remain nearly the same when researchers administer the instrument at different occasions. Also, scores need to be consistent. When an individual answer certain questions one way, the individual should consistently answer closely related questions in the same way. Validity is the development of sound evidence to demonstrate that the test interpretation (of scores about the concept or construct that the test is assumed to measure) matches its proposed use (AERA, APA, NCME, 1999).

On a theoretical level, reliability is concerned with the effect of error on the consistency of scores. We must be concerned about errors of measurement, or unreliability, because we know that test performance is affected by factors other than the abilities we want to measure. Bachman (1990) groups factors other than communicative language ability that affect performance on language tests into the following three broad categories: (1) test method facets; (2) attributes of the test taker which are not among the language abilities we are interested in measuring; and (3) random factors which are not stable over time and cannot be predicted. Systematicity is a feature of the test method facets which means that they do not vary from one test administration to the next. Attributes of the test takers refer to individual characteristics such as cognitive style and content knowledge of a particular areas, or group characteristics such as gender, race, and ethnic background. Similar to test method facets, such attributes are systematic to the extent that they affect individuals' test performance at a regular pattern. An individual's test score is not affected only by systematic sources of error. Unsystematic, or random factors also exert their impact to some extent.

Ary et al. (2010) introduce three sources of random error that may lead to inconsistency in scores:

1. Characteristics of the individual: variations in individuals' motivation, level of fatigue, physical health, anxiety, and other mental and emotional factors may affect test results.

2. The administration procedures and conditions: administering or scoring of a test may depart from standardized procedures. Testing conditions such as light, heat, ventilation, time of day, and the presence of distractions can influence test performance. Also, test-taking instructions and directions may not be clear enough. The scoring method may introduce a further source of error.

3. The testing instrument: a major threat to reliability is a test being brief. The longer the test, the more reliable it is likely to be. A short test gives a small sample of behavior and this may result in an unstable score. Luck has a greater chance to contribute in a short test than in a long test.

The effect of such random errors of measurement on the consistency of test scores is what reliability deals with. The classical true score theory approaches the issue of estimating reliability in three different ways, each of which is concerned with different sources of error. Error sources originated from within the test and scoring procedures are addressed in internal consistency estimates while stability estimates show the degree to which test scores are consistent across different administrations. The comparability of scores on alternate forms of a test is examined through the equivalence estimates (Bachman, 1990).

The equivalence estimates (parallel forms reliability) checks the equivalence of scores from alternate versions of a test. In cases where internal consistency estimates are not possible or suitable, the equivalence estimates are appropriate

options for estimating reliability. In some situations, equivalent forms of a test may already be used to decrease the practice effect or the chance of cheating. The parallel forms reliability is particularly suitable in such cases (Bachman, 1990). If the students' scores obtained from alternate forms of a test in different administrations are correlated, the related coefficient is known as the coefficient of stability and equivalence. Two facets of test reliability are revealed by this coefficient: fluctuations in test performance from one occasion to another and fluctuations from one form of the test to another. A high coefficient of stability and equivalence suggests that the same ability is measured by the two forms of the test and this measurement is consistent during different occasions. This is the most challenging and the most accurate measure available for estimating the reliability of a test (Ary et al. 2010).

In order to determine the reliability of alternate forms of a given test, the procedure used is to administer both forms to a group of individuals. The means and standard deviations for each of the two forms can then be computed and compared to determine their equivalence, after which the correlation between the two sets of scores can be computed. This correlation is then interpreted as an indicator of the equivalence of the two tests, or as an estimate of the reliability of either one (Bachman, 1990).

2.2 The Vocabulary Levels Test

The Vocabulary Levels Test was developed in the early 1980s by Paul Nation at the Victoria University of Wellington in New Zealand. Initially, it constituted a simple instrument serving the classroom purpose of helping teachers prepare a vocabulary teaching and learning plan. It was then published in Nation (1983, 1990) and has been extensively used in New Zealand and elsewhere since then (Xing & Fulcher, 2007). In an initial validation study in 1988, Read established the reliability of the instrument. He also found that individual scores for each frequency level formed an implicational scale according to which knowledge of lower-frequency words indicated knowing higher-frequency ones. Read's (1988) study was the last attempt for some time to validate the Levels Test. However, as Nation's book turned out to be a major vocabulary reference source, the test gained an international popularity.

Ten years after the test was first published, Norbert Schmitt revised the Levels Test in Nation's book (Version A) and wrote three additional versions (Versions B, C and D) using new bunches of words for each level. The original specifications remained intact in the new versions. No validation study was conducted for the versions written by Schmitt at that time. Still, the four versions' potential as a useful assessment instrument received considerable attention in numerous educational environments. Some vocabulary research studies have also used the tests as their instrument (e.g., Cobb, 1997; Schmitt and Meara, 1997; Laufer and Paribakht, 1998).

Beglar and Hunt (1999) administered two forms of Schmitt's (1993) version for the 2000- Word-Level and for the University-Word-Level to EFL learners in secondary and tertiary institutions in Japan. They used the results to select 54 items among the best-performing ones to develop two fresh tests for each level containing 27 items each. They then equated the two pairs of tests statistically. Schmitt, Schmitt, and Clapham (2001) undertook a similar test-development project with the four full forms of the test. They administered the tests to106 non-native speaking British university students and created two longer versions which included 30 items instead of the original 18. This article examines these two new versions developed by Schmitt et al. (2001).

The Vocabulary Levels Test used word–definition matching format to require test-takers to match the words to the definitions. Rather than giving a single estimate of total vocabulary size, it measures knowledge of words at five levels: 2000, 3000, 5000, 10,000, and academic English words. Each level contains 30 items which are arranged in 10 clusters. The ratio of different word classes in English is maintained in the test, with each section containing five noun clusters, three verb clusters, and two adjective clusters. The following illustrates the format of a noun cluster:

You must choose the right word to go with each meaning. Write the number of that word next to its meaning.

1 apparatus

2 compliment   _____ expression of admiration

3 ledge   _____ set of instruments or machinery

4 revenue   _____ money received by the government

5 scrap

6 tile

There are three definitions on the right and six words on the left. Candidates need to choose three out of the six words to match the three on the right. In total at each level, 30 definitions need to be matched to 30 out of 60 words. Schmitt et al. (2001) summarize the considerations kept in mind while writing each cluster as follows:

1. The options in this format are words instead of definitions.

2. The definitions are kept short, so that there is a minimum of reading, allowing for more items to be taken within a given period of time.

3. Words are learned incrementally, and tests should aim to tap into partial lexical knowledge. The Levels Test was designed to do this. The option words in each cluster are chosen so that they have very different meanings. Thus, even if learners have only a minimal impression of a target word's meaning, they should be able to make the correct match.

4. The clusters are designed to minimize aids to guessing. The target words are in alphabetical order, and the definitions are in order of length. In addition, the target words to be defined were selected randomly.

5. The words used in the definitions are always more frequent than the target words. The 2000 level words are defined with 1000 level words and, wherever possible, the target words at other levels are defined with words from the GSL (essentially the 2000 level). This is obviously important as it is necessary to ensure that the ability to demonstrate knowledge of the target words is not compromised by a lack of knowledge of the defining words.

6. The word counts from which the target words were sampled typically give base forms. However, derived forms are sometimes the most frequent members of a word family. Therefore, the frequency of the members of each target word family was checked, and the most frequent one attached to the test.

7. As much as possible, target words in each cluster begin with different letters and do not have similar orthographic forms. Likewise, similarities between the target words and words in their respective definitions were avoided whenever possible (pp. 59).

As was mentioned earlier, despite the widespread use of the Vocabulary Levels Test, very few studies have examined the reliability and validity of the test, none of which have been conducted in Iran. Also, since reliability will be a function not only of the test, but of the performance of the individuals who take the test, any given estimate of reliability based on the CTS model is limited to the sample of test scores upon which it is based (Bachman, 1990; Ary et al., 2010). Thus, the reliability estimates based on the CTS model reported in other studies cannot be transferred to the Iranian context, and we must always estimate the reliability of scores of the specific groups with whom we may want to use the test. This purpose of the current study is to estimate the equivalence and reliability of the two new versions of the Vocabulary Levels Test developed by Schmitt et al. (2001) for Iranian intermediate learners of English as a foreign language.

3. Research Questions

This article is an attempt to address the following two research questions formulated based on the purpose of the study:

RQ 1: Are version 1 and version 2 of the Vocabulary Levels Test parallel forms?

RQ 2: Are version 1 and version 2 of the Vocabulary Levels Test reliable tests?

4. Method

4.1 Participants

Seventy-five intermediate students who were studying English as a foreign language at the Iran Language Institute in Boukan served as the participants of the present study. Their proficiency level was judged to be intermediate based on the institute's placement test. The participants ranged in age from 16 to 27. Out of the 75 participants in this study, 46 were males and 29 females.

4.2 Instrumentation

Version 1 and version 2 of the Vocabulary Levels Test developed by Schmitt et al. (2001) were used as the data elicitation instruments. Each version is composed of five sections: 2000, 3000, 5000, 10000, and the academic vocabulary. Each section is made up of ten three-item clusters. The total possible score for each section is 30, and the total possible score for the whole test is 150 (see the Appendix).

4.3 Data Collection Procedure

Before the study was conducted, version 1 of the test was piloted on four intermediate students at the same institute from which the participants came from. The time needed to complete the test and difficulties they had in the process were observed, and the data was used to set the time and administration condition for the main study. The participants then took both versions of the test separately with one week interval between the two administrations. Half of the participants took version 1 first, and the other half took version 2 first to counterbalance the practice effect that would confound the relative equivalence of the two versions. Each participant took the two versions at the same testing conditions, including familiarity of the place, personnel, time of testing, and physical conditions.

4.4 Data Analysis

The data obtained from the participants was analyzed using version 20.0 of the Statistical Package for Social Science (SPSS). To examine the equivalence and reliability of the two versions and the individual sections, the means and standard deviations of the two versions and the individual sections were subjected to paired-sample t-tests and one-way ANOVAs, respectively. Furthermore, Pearson Product Moment correlation was used to examine the reliability of the two versions and the individual bands.

5. Results and Discussion

The first research question concerns the equality of the two versions of the test. As was mentioned earlier, two test are considered equal or parallel if their means and variances are the same (Bachman, 1990). Table 1 shows the descriptive statistics for the scores obtained from the two versions of the test.

Table 1. Descriptive Statistics for Version 1 and Version 2 Scores

	N	Minimum	Maximum	Mean	SD
Version 1	75	17.00	129.00	57.26	23.22
Version 2	75	15.00	131.00	59.40	22.77

The mean of the scores obtained from version 1 of the test is 57.26, while the mean of the scores obtained from version 2 is 59.40. The standard deviations for version and version 2 are 23.22 and 22.77, respectively. To determine whether the difference between the means of the two versions are significant, a paired-sample t-test was used. The results are shown in Table 2.

Table 2. Paired-samples T-Test Results for Version 1 and Version 2 Means

	Mean	SD	SEM	t	df	Sig.
V1 and V2	-2.13	8.09	0.93	-2.28	74	.025

The results indicate that the 2.13 difference between the means of the two versions is statistically significant (t = 2.28, N = 75, p< .05). Therefore, the means of the two versions are not the same. To compare the variances of the two versions of the test, their standard deviations were subjected to a one-way ANOVA. The results are shown in Table 3.

Table 3. One-Way ANOVA Results for Version 1 and Version 2 Standard Deviations

	Sum of squares	df	Mean square	F	Sig.
Between groups	170.66	1	170.66	0.323	0.571
Within groups	78302.66	148	529.07		
Total	78473.33	149			

The F value for the two standard deviations is .323 which is not significant, p= .571. This shows that the variances of the two versions are the same. However, since the means of the two versions are not the same, it is concluded that the two versions are not equal or parallel forms. Version 1 seems to be more difficult that version 2. To investigate the roots of this inequality between the two versions, the individual sections comprising the two versions were compared in terms of their means and standard deviations to determine if they are equal. Table 4 shows the descriptive statistics for the individual sections of each version.

Table 4. Descriptive Statistics for the Scores of the Individual Sections of Version 1 and Version 2

	Version 1		Version 2	
Section	Mean	SD	Mean	SD
2000	21.02	6.03	20.76	6.49
3000	12.52	5.76	14.18	5.90
5000	8.96	5.62	7.98	5.00
10000	2.09	2.69	1.81	2.96
Academic	12.66	6.73	14.65	6.63

To determine whether the differences between the means and variances of the individual sections are significant, paired-sample t-tests and one-way ANOVAs were used for the means differences and standard deviations differences, respectively. The results are shown in Table 5.

Table 5. Paired-Sample T-Tests and One-Way ANOVAs Results for the Means and Variances of the Individual Sections of Version 1 and Version 2

	Means			Variance		
Sections	Means diff.	t value	Sig.	SDs diff.	F value	Sig.
2000	0.26	.664	.509	0.46	.068	.795
3000	1.66	3.376	.001**	0.14	.000	1.000
5000	0.98	2.637	.010**	0.62	1.253	.265
10000	0.28	1.104	.273	0.27	.051	.822
Academic	1.99	4.382	.000**	0.1	3.311	.071

** Difference is significant at the .01 level.

As it can be seen in Table 5, some individual sections of the two versions are not equal or parallel. These sections are the 3000, 5000, and the academic sections whose means for the two versions are statistically different. Therefore, it is safe to claim that the inequality of the two versions as a whole results from the inequality of the 3000, 5000, and the academic sections. The inequality in each section may stem from the unbalanced distribution of items with high levels of item difficulty.

To address the second research question, which concerns estimation of the reliability of the two versions of the test through a parallel forms approach, the correlation between the two versions were calculated. Correlations tell us to what extent two variables are related or, in other words, related to each other. If there is a correlation between two variables, we should be able to impose a line on the scatterplot data points. The more tightly clustered around the line the data is, the stronger the correlation. Thus, the first assumption we must satisfy in order to test for correlation is that the relationship between the data is linear. The first step in performing correlation is to take a graphic look at our data (Larson-Hall, 2010). Figure 1 shows the scatterplot for the means of the scores obtained from the two versions.
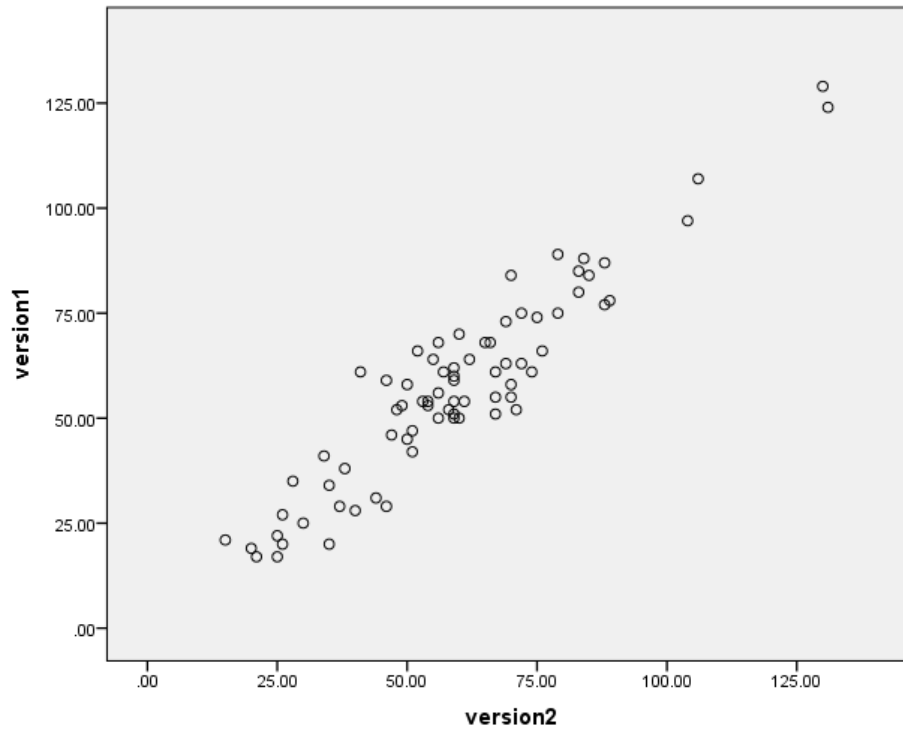

Figure 1. Scatterplot for the Means of Scores obtained from Version 1 and Version 2

As can be seen, although the points do not lie in a perfect line, there is an obvious upward trend in the presented data. It is therefore appropriate to test for a linear relationship in the data by performing a correlation. To estimate the correlation between the means of the two versions of the Vocabulary Levels Test, Pearson Product Moment correlation was used. The results are shown in Table 6.

Table 6. Pearson Product Moment Correlation between Version 1 and Version 2

Variables	Test	r	p	N
Version 1 and Version 2	2-tailed	.938**	.000	75

** Correlation is significant at the .01 level.

The correlation coefficient obtained for the two versions is significant beyond the .01 level (r = .938, p = .000, N = 75). This high level of correlation can be taken as evidence for the reliability of both version 1 and version 2 of the Vocabulary Levels Test. Additional support for the reliability of the two versions comes from the correlation between the individual sections. Table 7 shows the correlation coefficient obtained for the individual sections of the two versions.

Table 7. Pearson Product Moment Correlation between Individual Sections of Version 1 and Version 2

Section	r	p	N
2000	.849**	.000	75
3000	.732**	.000	75
5000	.825**	.000	75
10000	.702**	.000	75
Academic	.828**	.000	75

** Correlation is significant at the .01 level.

As is reported in Table 7, all the sections of the two versions are correlated with each other, which can be taken as additional support for the reliability of the two versions. Therefore, it can be concluded that version 1 and version 2 of the Vocabulary Levels Test are reliable tests.

6. Conclusion

The study reported in this paper aimed at examining the equivalence and reliability of the two new versions of the Vocabulary Levels Test for the Iranian learners of English as a foreign language. The results of the data analysis revealed high correlation between version 1 and version 2 and their individual sections. The two versions, therefore, are considered as highly reliable. However, the two versions were found not to be equivalent or parallel. Keeping this result in mind, treating the two versions as equal forms is untenable, hence vocabulary researchers are warned against using them as parallel tests particularly in the case of longitudinal or gain score research studies. However, given the relatively small scale of the differences between the mean scores that the two versions and their individual sections yield, they can probably be used in programs as alternate forms, as long as no high-stakes conclusions are drawn from a comparison between the two forms, and as long as the potential differences in scores between the two versions are kept in mind.

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency.* London: Continuum.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.

Bachman, L.F. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

Beglar, D. & Hunt, A. (1999). Revising and validating the 2000 word level and university word level tests. Language Testing, 16, 131-162.https://doi.org/10.1191/026553299666419728

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*, 301-315. https://doi.org/10.1016/S0346-251X(97)00024-9

D'Anna, C.A., Zechmeister, E.B., & Hall J.W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior, 23*(1), 109-122.https://doi.org/10.1080/10862969109547729

Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*(2), 174-187.

https://doi.org/10.2307/747863

Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary from oral input: A review article. *Applied Language Learning, 5*(1), 1-32.

Goulden, R., Nation, P., & Read J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*(4), 341-363.https://doi.org/10.1093/applin/11.4.341

Haynes, M., & Baker, I. (1993). American and Chinese readers learning from lexical familiarization in English texts. In T. Huckin, M. Haynes, & J. Coady (Eds.). *Second language reading and vocabulary acquisition* (pp. 130-152). Norwood, NJ: Ablex.

Hinkel, E. (2001). Giving examples and telling stories in academic essays. *Issues in Applied Linguistics, 12,* 149-170.

Hu, M., & Nation, I.S.P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 23*(1), 403-430.

Huckin, T. & Bloch, J. (1993). Strategies for inferring word meaning in contexts: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.). *Second language reading and vocabulary acquisition* (pp. 130-152). Norwood, NJ: Ablex.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*(3), 357-377. https://doi.org/10.1093/applin/19.3.357

Larson-Hall, J. (2010). A guide to doing statistics in second language research using SPSS. New York, NY: Routledge.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Lingui*stics, *16*(1), 307-322. https://doi.org/10.1093/applin/16.3.307

Laufer, B., & Paribakht, T.S. (1998). The relationship between passive and active vocabularies: effects of language learning context. *Language Learning, 48*, 365-391. https://doi.org/10.1111/0023-8333.00046

Lee, S. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System, 31,* 537-561.https://doi.org/10.1016/j.system.2003.02.004

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.). Performance and competence in second language acquisition (pp. 35-53). Cambridge: Cambridge University Press.

Nation, P. (1983). Testing and teaching vocabulary. *Guidelines, 5,* 12-25.

Nation, I.S.P. (1990). *Teaching and learning vocabulary.* New York: Newbury House.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63(*1) 59-82. https://doi.org/10.3138/cmlr.63.1.59

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal, 19,* 77-89. https://doi.org/10.1177/003368828801900202

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18* (1), 55-88. https://doi.org/10.1191/026553201668475857

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* London: Palgrave Macmillan. https://doi.org/10.1057/9780230293977

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition, 19,* 17-36. https://doi.org/10.1017/s0272263197001022

Wilkins, D.A. (1972). Linguistics in language teaching. London: Arnold.

Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System 35,* 182–191. https://doi.org/10.1016/j.system.2006.12.009.

Appendices

The Vocabulary Levels test (versions 1 and 2)

This is a vocabulary test.   You must choose the right word to go with each meaning. Write the number of that word next to its meaning.   Here is an example.

1 business
2 clock   _____ part of a house
3 horse   _____ animal with four legs
4 pencil   _____ something used for writing
5 shoe
6 wall

You answer it in the following way.

1 business
2 clock   ___6__ part of a house
3 horse   ___3__ animal with four legs
4 pencil   ___4__ something used for writing
5 shoe
6 wall

Some words are in the test to make it more difficult.   You do not have to find a meaning for these words.   In the example above, these words are business, clock, and shoe.

If you have no idea about the meaning of a word, do not guess. But if you think you might know the meaning, then you should try to find the answer.

Version 1   The 2,000 word level

1 birth
2 dust   _____ game
3 operation   _____ winning
4 row   _____ being born
5 sport
6 victory

1 choice
2 crop   _____ heat
3 flesh   _____ meat
4 salary   _____ money paid regularly for
5 secret   doing a job
6 temperature

1 cap
2 education   _____ teaching and learning
3 journey   _____ numbers to measure with
4 parent   _____ going to a far place
5 scale
6 trick

1 attack
2 charm   _____ gold and silver
3 lack   _____ pleasing quality
4 pen   _____ not having something
5 shadow
6 treasure

1 cream
2 factory   _____ part of milk
3 nail   _____ a lot of money
4 pupil   _____ person who is studying
5 sacrifice
6 wealth

1 adopt
2 climb   _____ go up
3 examine   _____ look at closely
4 pour   _____ be on every side
5 satisfy
6 surround

1 bake
2 connect   _____ join together
3 inquire   _____ walk without purpose
4 limit   _____ keep within a certain size
5 recognize
6 wander

1 burst
2 concern   _____ break open
3 deliver   _____ make better
4 fold   _____ take something to someone
5 improve
6 urge

1 original
2 private   _____ first
3 royal   _____ not public
4 slow   _____ all added together
5 sorry
6 total

1 brave
2 electric   _____ commonly done
3 firm   _____ wanting food
4 hungry   _____ having no fear
5 local
6 usual

Version 1   The 5,000 word level

1 balloon
2 federation   _____ bucket
3 novelty   _____ unusual interesting thing
4 pail   _____ rubber bag that is filled
5 veteran   with air
6 ward

1 alcohol
2 apron   _____ stage of development
3 hip   _____ state of untidiness or
4 lure   dirtiness
5 mess   _____ cloth worn in front to
6 phase   protect your clothes

1 apparatus
2 compliment   _____ expression of admiration
3 ledge   _____ set of instruments or
4 revenue   machinery
5 scrap   _____ money received by the
6 tile   Government

1 bulb
2 document   _____ female horse
3 legion   _____ large group of soldiers or
4 mare   people
5 pulse   _____ a paper that provides
6 tub   information

1 concrete
2 era   _____ circular shape
3 fiber   _____ top of a mountain
4 loop   _____ a long period of time
5 plank
6 summit

1 blend
2 devise   _____ mix together
3 hug   _____ plan or invent
4 lease   _____ hold tightly in your arms
5 plague
6 reject

1 abolish

2 drip   _____ bring to an end by law

3 insert   _____ guess about the future

4 predict   _____ calm or comfort someone

5 soothe

6 thrive

1 bleed

2 collapse   _____ come before

3 precede   _____ fall down suddenly

4 reject   _____ move with quick steps and

5 skip   jumps

6 tease

1 casual

2 desolate   _____ sweet-smelling

3 fragrant   _____ only one of its kind

4 radical   ____ good for your health

5 unique

6 wholesome

1 gloomy

2 gross   _____ empty

3 infinite   _____ dark or sad

4 limp   _____ without end

5 slim

6 vacant

Version 1   The 10,000 word level

1 antics

2 batch   _____ foolish behavior

3 connoisseur   _____ a group of things

4 foreboding   _____ person with a good

5 haunch   knowledge of art or music

6 scaffold

1 auspices

2 dregs   _____ confused mixture

3 hostage   _____ natural liquid present in the

4 jumble   mouth

5 saliva   _____ worst and most useless

6 truce   parts of anything

1 casualty

2 flurry   _____ someone killed or injured

3 froth   _____ being away from other

4 revelry   people

5 rut   _____ noisy and happy

6 seclusion   celebration

1 apparition

2 botany   _____ ghost

3 expulsion   _____ study of plants

4 insolence   _____ small pool of water

5 leash

6 puddle

1 arsenal

2 barracks   _____ happiness

3 deacon   _____ difficult situation

4 felicity   _____ minister in a church

5 predicament

6 spore

1 acquiesce

2 bask   _____ to accept without protest

3 crease   _____ sit or lie enjoying warmth

4 demolish   _____ make a fold on cloth or

5 overhaul   paper

6 rape

1 blaspheme

2 endorse   _____ slip or slide

3 nurture   _____ give care and food to

4 skid   _____ speak badly about God

5 squint

6 straggle

1 clinch

2 jot   _____ move very fast

3 mutilate   _____ injure or damage

4 smolder   _____ burn slowly without flame

5 topple

6 whiz

1 auxiliary

2 candid   _____ bad tempered

3 luscious   _____ full of self importance

4 morose   _____ helping, adding support

5 pallid

6 pompous

1 dubious

2 impudent   _____ rude

3 languid   _____ very ancient

4 motley   _____ of many different kinds

5 opaque

6 primeval

Version 1   Academic Vocabulary

1 benefit

2 labor   _____ work

3 percent   _____ part of 100

4 principle   _____ general idea used to

5 source   guide one's actions

6 survey

1 element   _____ money for a special

2 fund   purpose

3 layer   _____ skilled way of doing

4 philosophy   something

5 proportion   _____ study of the meaning

6 technique   of life

1 consent

2 enforcement   _____ total

3 investigation   _____ agreement or permission

4 parameter   _____ trying to find information

5 sum   about something

6 trend

1 decade

2 fee   _____ 10 years

3 file   _____ subject of a discussion

4 incidence   _____ money paid for services

5 perspective

6 topic

1 colleague

2 erosion   _____ action against the law

3 format   _____ wearing away gradually

4 inclination   _____ shape or size of something

5 panel

6 violation

1 achieve

2 conceive   _____ change

3 grant   _____ connect together

4 link   _____ finish successfully

5 modify

6 offset

1 convert

2 design   _____ keep out

3 exclude   _____ stay alive

4 facilitate   _____ change from one thing

5 indicate   into another

6 survive

1 anticipate

2 compile   _____ control something skillfully

3 convince   _____ expect something will

4 denote   happen

5 manipulate   _____ produce books and

6 publish   newspapers

1 equivalent

2 financial   _____ most important

3 forthcoming   _____ concerning sight

4 primary   _____ concerning money

5 random

6 visual

1 alternative

2 ambiguous   _____ last or most important

3 empirical   _____ something different that

4 ethnic   can be chosen

5 mutual   _____ concerning people from

6 ultimate   a certain nation

Version 2   The 2,000 word level

1 copy
2 event   _____ end or highest point
3 motor   _____ this moves a car
4 pity   _____ thing made to be like
5 profit   another
6 tip

1 accident
2 debt   _____ loud deep sound
3 fortune   _____ something you must pay
4 pride   _____ having a high opinion of
5 roar   yourself
6 thread

1 coffee
2 disease   _____ money for work
3 justice   _____ a piece of clothing
4 skirt   _____ using the law in the right
5 stage   way
6 wage

1 clerk
2 frame   _____ a drink
3 noise   _____ office worker
4 respect   _____ unwanted sound
5 theater
6 wine

1 dozen
2 empire   _____ chance
3 gift   _____ twelve
4 opportunity   _____ money paid to the
5 relief   government
6 tax

1 admire
2 complain   _____ make wider or longer
3 fix   _____ bring in for the first time
4 hire   _____ have a high opinion of
5 introduce   someone
6 stretch

1 arrange
2 develop   _____ grow
3 lean   _____ put in order
4 owe   _____ like more than something
5 prefer   else
6 seize

1 blame
2 elect   _____ make
3 jump   _____ choose by voting
4 manufacture   _____ become like water
5 melt
6 threaten

1 ancient
2 curious   _____ not easy
3 difficult   _____ very old
4 entire   _____ related to God
5 holy
6 social

1 bitter
2 independent   _____ beautiful
3 lovely   _____ small
4 merry   _____ liked by many people
5 popular
6 slight

Version 2   The 3,000 word level

1 bull
2 champion   _____ formal and serious manner
3 dignity   _____ winner of a sporting event
4 hell   _____ building where valuable
5 museum   objects are shown
6 solution

1 blanket
2 contest   _____ holiday
3 generation   _____ good quality
4 merit   _____ wool covering used on
5 plot   beds
6 vacation

1 comment
2 gown   _____ long formal dress
3 import   _____ goods from a foreign
4 nerve   country
5 pasture   _____ part of the body which
6 tradition   carries feeling

1 administration
2 angel   _____ group of animals
3 frost   _____ spirit who serves God
4 herd   _____ managing business and
5 fort   affairs
6 pond

1 atmosphere
2 counsel   _____ advice
3 factor   _____ a place covered with grass
4 hen   _____ female chicken
5 lawn
6 muscle

1 abandon
2 dwell   _____ live in a place
3 oblige   _____ follow in order to catch
4 pursue   _____ leave something
5 quote   permanently
6 resolve

1 assemble
2 attach   _____ look closely
3 peer   _____ stop doing something
4 quit   _____ cry out loudly in fear
5 scream
6 toss

1 drift
2 endure   _____ suffer patiently
3 grasp   _____ join wool threads together
4 knit   _____ hold firmly with your hands
5 register
6 tumble

1 brilliant	
2 distinct	_____ thin
3 magic	_____ steady
4 naked	_____ without clothes
5 slender	
6 stable	

1 aware	
2 blank	_____ usual
3 desperate	_____ best or most important
4 normal	_____ knowing what is happening
5 striking	
6 supreme	

Version 2   The 5,000 word level

1 analysis	
2 curb	_____ eagerness
3 gravel	_____ loan to buy a house
4 mortgage	_____ small stones mixed with
5 scar	sand
6 zeal	

1 cavalry	
2 eve	_____ small hill
3 ham	_____ day or night before a
4 mound	holiday
5 steak	_____ soldiers who fight from
6 switch	horses

1 circus	
2 jungle	_____ musical instrument
3 nomination	_____ seat without a back or
4 sermon	arms
5 stool	_____ speech given by a priest in
6 trumpet	a church

1 artillery	
2 creed	_____ a kind of tree
3 hydrogen	_____ system of belief
4 maple	_____ large gun on wheels
5 pork	
6 streak	

1 chart	
2 forge	_____ map
3 mansion	_____ large beautiful house
4 outfit	_____ place where metals are
5 sample	made and shaped
6 volunteer	

1 contemplate	
2 extract	_____ think about deeply
3 gamble	_____ bring back to health
4 launch	_____ make someone angry
5 provoke	
6 revive	

1 demonstrate	
2 embarrass	_____ have a rest
3 heave	_____ break suddenly into small
4 obscure	pieces
5 relax	_____ make someone feel shy or
6 shatter	nervous

1 correspond	
2 embroider	_____ exchange letters
3 lurk	_____ hide and wait for someone
4 penetrate	_____ feel angry about something
5 prescribe	
6 resent	

1 decent	
2 frail	_____ weak
3 harsh	_____ concerning a city
4 incredible	_____ difficult to believe
5 municipal	
6 specific	

1 adequate	
2 internal	_____ enough
3 mature	_____ fully grown
4 profound	_____ alone away from other
5 solitary	things
6 tragic	

Version 2   The 10,000 word level

1 alabaster	
2 chandelier	_____ small barrel
3 dogma	_____ soft white stone
4 keg	_____ tool for shaping wood
5 rasp	
6 tentacle	

1 benevolence	
2 convoy	_____ kindness
3 lien	_____ set of musical notes
4 octave	_____ speed control for an
5 stint	engine
6 throttle	

1 bourgeois	
2 brocade	_____ middle class people
3 consonant	_____ row or level of something
4 prelude	_____ cloth with a pattern or gold
5 stupor	or silver threads
6 tier	

1 alcove	
2 impetus	_____ priest
3 maggot	_____ release from prison early
4 parole	_____ medicine to put on wounds
5 salve	
6 vicar	

1 alkali	
2 banter	_____ light joking talk
3 coop	_____ a rank of British nobility
4 mosaic	_____ picture made of small pieces
5 stealth	of glass or stone
6 viscount	

1 dissipate	
2 flaunt	_____ steal
3 impede	_____ scatter or vanish
4 loot	_____ twist the body about
5 squirm	uncomfortably
6 vie	

1 contaminate
2 cringe   _____ write carelessly
3 immerse   _____ move back because of fear   1 alter
4 peek   _____ put something under water   2 coincide   _____ change
5 relay   3 deny   _____ say something is not true
6 scrawl   4 devote   _____ describe clearly and exactly
                                                           5 release
1 blurt   6 specify
2 dabble   _____ walk in a proud way
3 dent   _____ kill by squeezing someone's   1 correspond
4 pacify   throat   2 diminish   _____ keep
5 strangle   _____ say suddenly without   3 emerge   _____ match or be in agreement
6 swagger   thinking   4 highlight   with
                                                           5 invoke   _____ give special attention
1 illicit   6 retain   to something
2 lewd   _____ immense
3 mammoth   _____ against the law   1 bond
4 slick   _____ wanting revenge   2 channel   _____ make smaller
5 temporal   3 estimate   _____ guess the number or size
6 vindictive   4 identify   of something
                                                           5 mediate   _____ recognizing and naming
1 indolent   6 minimize   a person or thing
2 nocturnal   _____ lazy
3 obsolete   _____ no longer used   1 explicit
4 torrid   _____ clever and tricky   2 final   _____ last
5 translucent   3 negative   _____ stiff
6 wily   4 professional _____ meaning `no' or `not'
                                                           5 rigid
                                                           6 sole

Version 2   Academic Vocabulary
                                                           1 abstract
1 area   2 adjacent   _____ next to
2 contract   _____ written agreement   3 controversial _____ added to
3 definition   _____ way of doing something   4 global   _____ concerning the whole world
4 evidence   _____ reason for believing   5 neutral
5 method   something is or is not true   6 supplementary
6 role

1 debate
2 exposure   _____ plan
3 integration   _____ choice
4 option   _____ joining something into a
5 scheme   whole
6 stability

1 access
2 gender   _____ male or female
3 implementation _____ study of the mind
4 license   _____ entrance or way in
5 orientation
6 psychology

1 accumulation
2 edition   _____ collecting things over time
3 guarantee   _____ promise to repair a broken
4 media   product
5 motivation   _____ feeling a strong reason or
6 phenomenon   need to do something

1 adult
2 exploitation   _____ end
3 infrastructure _____ machine used to move
4 schedule   people or goods
5 termination   _____ list of things to do at
6 vehicle   certain times