TAV of Arabic Language Measurement

Harun Baharudin

Faculty of Education, National University of Malaysia (UKM)
Email: harunbaharudin@yahoo.com.my

Zawawi Ismail

Faculty of Education, University of Malaya
Email: zawawiismail@um.edu.my

Adelina Asmawi

Faculty of Education, University of Malaya
Email: adelina@um.edu.my

Normala Baharuddin

SPSR, Ministry of Education, Malaysia
Email: nbaharudin@gmail.com

Abstract

One of the many problems faced by second language students is the size of vocabulary. The aim of this study is to build a test to measure students' Arabic vocabulary size namely Test of Arabic Vocabulary (TAV). This research will then measure the validity and reliability of the test. The methodology of research involves the construction of tests, test items refinement process and the process of investigating test reliability and validity. Sample of research includes 33 pre university students of an Islamic secondary school in Temerloh, Pahang, Malaysia. This research is directed to follow an internal consistency test to observe the reliability of co-efficiency that is the correlation value of tests and retest, and the value of alpha Cronbach and Cohen Kappa. Results show that the value of correlation is high, that is .842, and the co-efficient of alpha Cronbach is high, .895, whereas the co-efficient of Cohen Kappa is also high, .95. Proof of content validity is reviewed by panel experts. The implication of this study is that the test can be used to measure TAV due to the test meeting criteria of reliable and valid psychometric characteristics.

Keywords: Arabic vocabulary test; validity; reliability;

1. Introduction

Vocabulary is an essential need in learning a second or foreign language. Students who are capable of mastering language skills are students who are able to master vocabulary at a fundamental stage of a targeted language (Hunt & Beglar, 2005; Tuaymah, 1998; Asmah 1984; Abd. Aziz, 2000). Based on the Kurikulum Bersepadu Sekolah Rendah (KBSR) (*Integrated Curriculum for Primary Schools*) and Kurikulum Bersepadu Sekolah Menengah (KBSM) (*Integrated Curriculum for Secondary Schools*) Arabic Language Curriculum Specifications, the aim for teaching and learning of Arabic language is to enable students to master Arabic well particularly in the aspect of vocabulary. This aim emphasises the aspect of size (value) and the appropriate use of vocabulary. However, the evaluation and measurement of the size of the Arabic vocabulary still be estimate based on mastery of language skills in general and implicitly (Che Radiah & Norhayuza, 2011; al-Batal, 2006; Ryding, 2006).

Although there are a number of studies on Arabic vocabulary size, the issue of the instrument to empirically measure students' Arabic vocabulary size is still underdeveloped. The issue of the total or size of the Arabic vocabulary requires an empirical assessment to measure or estimate the total number (size) of vocabulary that is mastered. This is important to ensure that the specific objective of vocabulary mastery that is set in the education syllabus, can be evaluated and measured clearly and precisely. The assessing of the Arabic vocabulary size in an empirical way is not given emphasis in Malaysia. Students' achievement is usually evaluated through their language proficiency thus if the

proficiency is high, it is assumed that the student had mastered Arabic vocabulary, and vice versa.

This study aims to explore students' perceptual learning style preferences as well as whether any linkages between language learning styles and such variables as field of study, study length, gender, age, language learning experience, and English proficiency level subsist.

2. Literature Review

Aspects of vocabulary size is one measure of language proficiency to mastering more amount vocabulary. In fact, most of the studies show that the level of vocabulary size correlated with language proficiency (Gu, 1998). Students will to absorb much language learning by mastering a large quantity of vocabulary. Students can easily get something new if knowing more words (Curtis, 2006). This is not surprising because the size of the vocabulary related with the ability to master the skills of reading, writing and public speaking skills and influence academic achievement (Saville-Troike, 1984; Laufer, 1997, Stehr, 2008). (Saville-Troike, 1984; Laufer, 1997, Stehr, 2008).

Some scholar suggests that the Arabic language students at the primary or beginning need to master the vocabulary size of 750 to 1000 words, in the middle level need to master of 1000 to 1500 words and at a high level need to master of 1500 and 2000 words (Tu 'aymah, 1986). Mat Taib (2006) also divide the vocabulary needs of three levels, namely 1000 to 1500 words at the beginning of 1500 to 2500 words to the second level and 2500 to 3500 words for the third stage. Al-Cancel (2006) also estimated that the total vocabulary needed to achieve high efficiency levels required by 3000 to 3500 words. Total enough vocabulary such as students need to make use of the dictionary and morphological knowledge in learning the Arabic vocabulary (Tu'aymah, 1986, al-Cancel, 2006).

Based on the discussion of vocabulary size in the view of some scholars and based on the objectives of the syllabus Arabic (can be seen in Table 1) , it can be concluded that students should master at least 3000 words during their studies at the school. Vocabulary size is quite reasonable in view of the long duration of their learning over from the primary to the secondary school level lasting more than 10 years.

Table 1. Vocabulary Size by the Ministry of Education Malaysia Syllabus

Level	Vocabulary Size
Arabic Language for Comunication	3000
Advance Arabic Language	2000
Arabic Language in KBSM syllabus	1500
Arabic language in jQAF (jawi, quran, Arabic & Fardu Ain) program	600

3. Objective of the Study

The objective of this study is to develop a test to measure students' mastery of Arabic vocabulary and to determine the psychometric properties of the test.

4. Research Questions

The research questions are:
1. To what extent does the test (constructed to determine Arabic vocabulary size) have psychometric reliability (a) test and retest (b) internal consistency of alpha Cronbach and (c) Cohen Kappa value?
2. To what extent does the test (constructed to determine Arabic vocabulary size) have psychometric properties of content validity?

5. Research Methodology

The methodology of research involves the process of constructing items for both tests as well as the process of refining both tests by item analysis. Additionally, proof of test reliability has been investigated by using the Cohen Kappa internal consistency method. The reliability refers to how far a test is able to consistently measure the element that is needed to be measured (Gay, Mills & Airasian 2006). The proof of test validity has also been investigated by content validation and concurrent validation. Content validity involves systemic investigation of test content to determine whether the content of test includes behaviour representation in the domain that is required to be measured (Gay, Mills & Airasian, 2006). In this

study, panel experts have been referred to for content validity.

6. Test Construction

The Arabic vocabulary size test in this study is more referred to as a proficiency test that is not based on any particular syllabus. According to Hasan Basri (2002), a language proficiency test is a test that may or may not be based on a particular syllabus. Furthermore, the test is also beneficial for the study of Islamic high school students in the future. Through the test of Arabic vocabulary size, students can be evaluated based on the estimation of their Arabic vocabulary size.

To measure vocabulary size, this research uses a procedure that is commonly used to measure and estimate the size of vocabulary. According to John (2000), a researcher usually uses a word sample that is taken from a complete dictionary to represent each word in an entry. This is because it is impossible to test each and every word. The following test will move to a number of subject groups and the result will be multiplied by the number of words to estimate the total size of vocabulary.

However, the use of dictionary to represent words is not viewed as appropriate and it is seen to have several disadvantages where too broad and common words are selected (Lorge & Chall, 1963). This is because a word entry in a dictionary typically has various meanings and uses (Nation, 1993). Moreover, using a dictionary to draw a sample of words has the disadvantage of involving technical aspects such as difficulty to accurately determine the estimated total number of vocabulary (Meara, 1996).

As this was the case, several researchers (Quinn, 1968; Harlech-Jones, 1983; Nurweni &Read, 1999) have used the frequency word list as the basis for their vocabulary sample size to measure students' vocabulary size. It is assumed better to draw word samples from a list of words that has been refined rather than from a dictionary.

Therefore, from the aspect of Arabic vocabulary size, the researcher extracted word samples from the Arabic frequency word list from a book entitled A Word Count of Modern Arabic Prose (Landau, 1959). This book lists specific Arabic words as many as 12,400 from a variety of Arabic prose and lists of words from daily newspapers in several Arab countries such as Egypt (Brill, 1940). The purpose of the frequency word list is to ensure that the items chosen for the test cover a wide range of words based on the frequency. The procedure of the analysis can ensure that the test items represent the entire field of content that is to be tested. This can enhance the content validity of a test.

Word items were selected from every 1000 to 4000 words. From every 1000 words, 50 words were randomly selected. These items consist of multiple words including nouns, verbs and articles. In total, 200 items were selected for the test based on each 1000 frequency level.

Subsequently, each word is evaluated based on its general usage by students in Malaysian schools. To refine this further, the researcher referred to a list of Arabic vocabulary to meet the suitability of South-East Asian students, as conducted by Abdul Rahman (1994). The lists of words are compiled based on five key evaluation criteria which are popular vocabulary, extent of use, questionnaire responses, essential vocabulary (Al Quran), and teaching and learning abilities (short words, borrowed words and clarity of meaning). This step is important to encourage word selection that is not too foreign in terms of exposure or aspect of learning.

In the following step, the researcher observed and examined words that have been used and learned from text books. For this purpose, Arabic communication text books and Arabic literature text books have been used. These text books are able to provide evidence and samples of vocabulary used throughout Arabic lessons in schools. The processes of selecting and evaluating vocabulary are as shown in Figure 1.
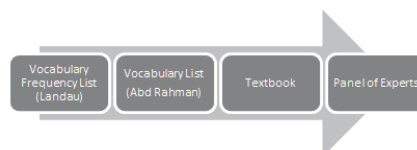

Figure 1. Word Selection Process to Measure Students' Arabic Vocabulary Size

For the purpose of refining the items, quantitative analysis was carried out to measure the items difficulty index (p) and items discrimination index (D) (Anastasi & Urbina, 1997). Item difficulty is defined as the percentage of people who correctly answer the item (Anastasi & Urbina, 1997). Whereas item discrimination refers to the effectiveness of an item to

discriminate between higher-scoring students and lower-scoring ones in a particular test (Aiken & Groth-Marnat, 2006). According to Anatasi & Urbina (1997), the best items are items that have a difficulty index in the middle range. Items with a difficulty of 0.15 to 0.85 are generally considered good to include in a test. Items with a discrimination index of 0.30 and above are considered acceptable to include in a test (Aiken & Groth-Marnat, 2006).

Based on findings for both indexes, 40 word items were selected to represent 4000 words. There are four parts to the test which represent four stages of word frequency that are, (1) 1000 words, (2) 2000 words, (3) 3000 words, and (4) 4000 words. The difficulty level of word increases with each stage. These words were arranged according to the available difficulty index. Table 2 shows the number of items that were selected for each stage based on results of item analyses.

Table 2. Items Selected Based On Results of Item Analyses

Arabic Voc. Level	No Of Items 0.15 < P > 0.85 And D > 0.3	Number After Item Refinement
1000	20	10
2000	14	10
3000	14	10
4000	13	10
Total	**61**	**40**

For the purpose of calculating test results, for every word that is answered correctly one mark is given. The total number of words that are answered correctly is the score of the test. According to Lorge & Chall (1963), if one was aware of the entire word sample for the first 1000, half of the second 1000 sample, and a quarter of the third 1000 sample, one can be estimated to have a vocabulary size of 1750 when all of it is totalled at each stage that is 1000 for the first stage, 500 for the second and 250 for the third.

To confirm that the review of the test answers is carried out properly, samples of answer sheets that have been examined are referred to three experts in the field for review. This is done to refine the students' answers that are predicted to be of various responses thus the review of student answers is only based on specific answers. The test scores are calculated based on percentage, then the mean score for each question that is tested as well as the mean scores of the entire test.

As for the design and format of the test, the researcher used bilingual test format. This format is used by researchers to measure vocabulary size (John, 2000) and among these researchers are Nation (2001) and Nurweni (1995), who conducted a study to measure English vocabulary size of Indonesian students. This format of the test serves to assist students in stating their understanding of targeted word without being challenged by the lack of knowledge of Arabic synonyms and phrases.

Following this format, students were required to provide definitions in Malay based on the given expression which were in the form of sentences, Quranic verses and phrases. According to Henning (1991), using sentences in the context or through language expressions may facilitate students' understanding of words. Additionally, details of test responses were constructed by students themselves who were not provided with multiple choice answers.

Additionally, the selected words did not include nouns such as names of people, places and others. This is because the selection of nouns is higher than the selection of verbs. According to Webb (2005), using a ratio of 6:4 nouns to verbs, it can be seen that nouns are more commonly used than verbs. This is supported by Kucera & Francis (1967) who stated that this rate is a comparable estimation of frequency in a situation. On top of that, this test also prevented students from focusing on grammatical errors. To conclude, the vocabulary test section of this study will measure Arabic vocabulary size of students.

7. Findings and Discussion

7.1 Reliability of Test

7.1.1 Test and retest

The test and retest approach in investigating the consistency of a test has been suggested by Baba (2002) and is seen as an approach that is generally accepted (Bachman, 1990). For this purpose, the test was carried out twice with a one week interval between them. According to Bachman (1990), the interval is to minimise any sign of practice and learning. Table 3 shows the scores for both tests

Table 3. *Scores of Test and Retest*

	N	Min.	Max.	Mean	SD
Test 1	30	0	22	6.26	5.91
Test 2	30	0	24	8.43	6.78

The results of correlation test scores 1 and retest 2 are significant $r= 0.842$ ($p < 0.0001$). This shows that the test reveals scores that are comparable for both of the tests. Hence, the test proves to have a high level of consistency. The strong correlation value between the scores of the two tests shows that the participants tend to give the same response in both tests.

7.1.2 Alpha Cronbach

Bachman (1990) suggested that it is required to carry out internal consistency to estimate the reliability of tests. For this reason, alpha Cronbach is used to achieve internal consistency that is more accurate. The alpha Cronbach value is 0.895 and this shows that the test for the vocabulary size has a high consistency thus can be accepted.

7.1.3 Cohen Kappa

The following consistency process is to obtain the Cohen Kappa scale. To do so, test questions were posed to ten students. Their answers were reviewed by 2 panels who awarded marks to each student. The analysis of agreement between the evaluator and two expert panels is presented in Table 4.

For the purpose of this agreement, the researcher used the Cohen Kappa scale to calculate the agreement. A number of local researchers have used the Cohen Kappa calculation method such as Nik Mohd Rahimi (2005), Azhar (2006), and Zawawi (2009). To perform the calculation, the following formula was used (Landis & Kosh, 1977; Zahrah, 2002):

Formula : $K= \dfrac{fa - fc}{N - fc}$

K= Coefficient of agreement
fa=Unit that is agreed
fc=Probable unit that is set at 50% from N
N=Total number of unit
The Kappa coefficient agreement that is suggested by Flesh (1981) is as shown in Table 4:

Table 4. Kappa Coefficient Agreement

Kappa value	Interpretation
Below 0.00	Poor
0.00 – 0.20	Unsatisfactory
0.21 – 0.40	Fairly Satisfactory
0.41 – 0.60	Satisfactory
0.61 – 0.80	Good
0.81 – 1.00	Excellent

The Kappa agreement value for Arabic vocabulary test can be seen in Table 5. This table shows that each of the Arabic vocabulary test data was handed to two panel experts.

Table 5. Kappa Agreement Value

Type of data	Panel 1	Panel 2	Average agreement	Coefficient of Cohen Kappa
Vocabulary test	K=$\dfrac{392-200}{400-200}$ K=$\dfrac{192}{200}$ K=0.96	K=$\dfrac{389-200}{400-200}$ K=$\dfrac{189}{200}$ K=0.945	$\dfrac{0.96+0.945}{2}$ = 0.95	0.95

For data consistency of Arabic vocabulary test, the researcher once again selected 30 test results to hand to two panel experts that were appointed. The 30 test results were randomly selected in accordance to Borg and Gall (1983) who state that the smallest unit for correlation analysis is 30 people. Correlation analysis on the other hand is conducted based on suggestions made by Hassan Basri (2002) and Mohd. Majid (2000). The score results from both scholars are compared with the scores that by the researcher.

Results from the correlation analysis shows the data consistency is at a high level with the correlation test value between researcher and panel 1 at 0.79 whereas with panel 2 is at 0.82. This value is obtained based on the deduction made by Devies (1971) as shown in Table 6:

Table 6. Interpretation of Coefficient

Correlation coefficient value (r)	Coefficient Interpretation
1.00	Perfect
0.70-0.99	Very high
0.50-0.69	Strong
0.30-0.49	Average
0.10-0.29	Low
0.01-0.09	Can be ignored

7.2 Validity of study content

The following test validity process involved discussions with teachers and experts of the Arabic language. They are three university lecturers, a language teacher and a successful Arabic school teacher. The main topic of discussion was focused on the selection of suitable vocabulary as well as the difficulty level of words that are included in Arabic vocabulary tests. In the context of language expression however, discussion was focused on the suitability of structure and style of language in order to meet respondent's level of competency.

A list of vocabulary is divided into four sections to represent words based on frequency level starting from level 1000 of frequent words through to level 4000. As many as 200 words have been selected and distributed to the panels to state their agreement on the selected words and a suitable level of difficulty for them to be placed in. Researchers themselves made appointments with each panel to receive responses, comments and advice to improve the content validity level of test instrument.

To summarise, the panel of evaluators acknowledge that the test is able to measure the aspect of content that is needed. This is according to the responses that are stated in several series of discussions. The views and opinions of the entire panel have been considered to improve the content of the test. After receiving an agreement on the whole item and content of the test, the following step is to conduct a preliminary study to assess instrument reliability through observing a number of important factors such as the period set for answering tests, doubtful test questions and the order or format of instrument used.

8. Discussions

The empirical results above show that the TAV has been produced and improved, and has also met the psychometric criteria of reliability and validity. A high reliability value shows that the test is able to consistently measure student's vocabulary size. Besides, content validity of the test has also been confirmed by panel experts. They have confirmed that the test is appropriate and sufficient to use to measure students' Arabic vocabulary size.

From observation, the test can be used to measure the vocabulary size of students and does not burden because the format of the test is bilingual. This format of the test serves to assist students in stating their understanding of targeted word without being challenged by the lack of knowledge of Arabic synonyms and phrases. At the same time, this format encourages students to recall words they have previously learned. In this case, Messick (1996) states that the element of 'recalling' during a test is an element that is useful in a test. The element of 'recalling' is used systematically in relation to the use of context of language expression - in the form of sentences and Quranic verses - to facilitate students' understanding of words.

This study adds to the collection of research on development and production of instruments to measure Arabic vocabulary size that is still inadequate particularly in the context of Arabic language learning. This is in contrast to the history of English vocabulary level tests that empirically began when Nation (1983) developed a diagnostic test for

teachers to identify students' vocabulary size (Norbert, Diane & Caroline, 2001). The test created may be used by researchers to measure student's vocabulary size for Arabic language, thus has pedagogical implications for the learning of Arabic language.

The test is included in the discrete test construct category (discrete) (John & Carol, 2001) which measures knowledge of basic vocabulary or the use of open construct. This test uses a simple test structure that only takes into account the fundamental knowledge of vocabulary and excludes knowledge of grammar and reading abilities. According to Lynne (2002), vocabulary learning begins with measurement of vocabulary size through word recognition before moving further into lexical knowledge such as spelling, collocation, grammar and semantics. A student who recognises a word knows the word (Lynne, 2002).

Moreover, vocabulary size estimated through the calculation of words recognized provides an early guide to one's lexical knowledge. In other words, knowledge of students' vocabulary size is a beneficial start. With this test, the measurement of students' mastery of Arabic vocabulary can be accurately conducted and students' proficiency can be estimated. A student who has a large vocabulary size is one who has acquired most of the semantics of any language (McCarthy, 1988).

Although there have been researchers who have conducted tests to identify the Arabic vocabulary size of students, many have used tests that were not empirical. For example, many local researchers use assessment tests that are based on text book syllabuses (Irma Martiny, 2012; Zahriah, 2011 and Rahim, 2009). Nevertheless, there are researchers who estimate vocabulary size based on students' test performance and language proficiency. If a student is a good student or has good language proficiency, it is assumed that her/his vocabulary size is high. This is in contrast with research in English language education that happens to be more advanced and possesses a number of instruments for vocabulary tests such as Vocabulary Levels Test, Lexical Frequency Profile, Vocabulary Knowledge Scale and many more.

In conclusion, there is empirical evidence that the developed test has psychometric properties of reliability and validity and can be used to measure students' Arabic vocabulary size.

References

Ab. Rahim, Ismail. (1994). Pengajaran bahasa Arab sebagai bahasa kedua: Permasalahan dan penyelesaian. In L. Muhammad Bukhari, Persuratan Islam. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Abdul Aziz, Abdul Talib (1996). Menguji Kemahiran Bahasa : Prinsip, Teknik dan Contoh. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Abdul Rahman, Chik. (1994). Ma'ayir al-tafakhum fi al-alfaz fi ta'lim al-Arabiyyah li al-mujtama'at al-Islamiyyah fi janub sharq Asiyya. (T. D. Falsafah, Ed.) Mesir: Universiti Iskandariyyah.

Aiken, L. R., & Groth-Marnat, G. (2006). Psychological testing and assesment. Boston, MA: Pearson Education Group Inc.

Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed ed.). New Jersey: Prentice-Hall.

Asmah Omar. (1984). Kaedah Pengajaran Bahasa. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Azhar, Muhammad., Ahmad Kilani, Mohamed., & Sulaiman Shakib, Mohd. Noor. (2003). Masalah penterjemahan Melayu-Arab di kalangan pelajar Melayu. Persidangan Penterjemahan Antarabangsa ke-9 (pp. 344-361). Kuala Lumpur: Universiti Pendidikan Sultan Idris.

Baba, K. (2002). test review: Lex30. Language Testing Update, 32, 68-71.

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

al-Batal, M. (2006). Playing with words: teaching vocabulary in the Arabic curriculum. In. K. Wahba, Z. Taha, & L. England, Handbook for Arabic language teaching profesionals in the 21st century (pp. 331-340). Mahwah,

Brill, M. (1940). The basic word list of the Arabic daily newspaper. Jerusalem: Hebrew University Prev.

Che' Pee, Che'Hanafi (1991). Pelajar sekolah menengah agama: Cabaran dan masa hadapan. Journal Pendidikan Islam, 3, 13-16.

Gay, L. R., & Airasian, P. (2003). Educational research compentencies for analysis and application. New Jersey: Prenctice Hall.

Harlech-Jones, B. (1983). ESL proficieency and a word-frequency count. English language Teaching Journal 37, 62-70.

Hasan Basri, Awang Mat Dahan (2002). Language testing: The construction and validation. Kuala Lumpur: University of Malaya Press.

Hunt, A., & Belglar, D. (2005). A framework for developing EFL reading vocabulary. Reading in a Foreign Language, 17, 1-31.

Irma Martiny, M. Yasim (2012). Hubungan strategi pembelajaran kosa kata dengan tahap penguasaan kosa kata pelajar. Bangi: Kertas Projek Sarjana Universiti Kebangsaan Malaysia.

John, R., & Carol, A. C. (2001). A framework for second language vocabulary assessment. language Testing, 18(1), 1-32.

Kementerian Pendidikan Malaysia (1994, April 23-24). Metodologi pengajaran Bahasa Arab di sekolah menengah. Kertas kerja seminar metodologi bahasa Arab peringkat kebangsaan.

Kementerian Pelajaran Malaysia (2002a). Sukatan pelajaran Bahasa Arab Komunikasi. Kuala Lumpur: Jabatan Pendidikan Islam dan Moral.

Kementerian Pelajaran Malaysia (2002b). Sukatan pelajaran Bahasa Arab Tinggi KBSM. Kuala Lumpur: Jabatan Pendidikan Islam dan Moral.

Landau, M. J. (1959). A word count of modern arabic prose. New York: American council of learned societis.

Lorge, I., & Chall, J. (1963). Estimating the size of vocabularies of children and adults: An analysis of metodological issues. Journal of Experimental Education, 32(2), 147-157.

Lynne, C. (2002). measuring vocabulary size in English as an additional language. Language Teaching Research, 6(2), 145-173.

McCarthy (1988). Vocabulary and Language Teaching. New York: Longman

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

Messick, S. (1996). Validity and washback in language testing. Language Testing, 13, 241-256.

Nation, P. (1993). Using dictionaries to estimate vocabulary size: essential but rarely followed procedures. Language testing 10, 27-40.

Nik Mohd Rahimi, Nik Yusoff. (2005). Penilaian Kemahiran Mendengar dalam kurikulum bahasa Arab Komunikasi di sekolah-sekolah agama kerajaan negeri. Universiti Kebangsaan Malaysia, Bangi: Tesis PhD.

Norbert, S., Diane, S., & Caroline, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. Language Testing, 18(1), 55-88.

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesia university student. English for Spefific Purpose 18, 161-175.

Quinn, G. (1968). The English Vocabulary of some Indonesian University entrants. Salatiga, Indonesia: IKIP Kristen Satya Watjana.

Rahim, Ramly (2009). Penguasaan Perbendaharaan Kata bahasa Arab dalam kalangan murid sekolah Kebangsaan Menerusi Program J-QAF: Satu Kajian Kes. Wacana Pendidikan Islam (peringkat Nusantara) siri ke-7 (pp. 579-588). Bangi: Universiti kebangsaan Malaysia.

Saifuddin, Hussin (2002). Penguasaan Perbendaharaan kata bahasa Arab di kalangan pelajar tingkatan tiga: Satu kajian kes di sekolah Menengah Kebangsaan (Agama) Negeri Melaka. Kuala Lumpur: Disertasi Sarjana, Universiti Malaya.

Seliger, H., & Shohamy, E. (1989). Second language research methods. Oxford: Oxford Univercity Press.

Sohair, Abdel Moneim Sery (1990). Satu tinjauan tentang masalah-masalah pengajaran mata pelajaran bahasa Arab di sekolah-sekolah menengah agama di Malaysia. Bangi: Projek Penyelidikan Sarjana; Universiti Kebangsaan Malaysia.

Tu'aymah, Rusydi Ahmad (1998). Manahij tadris al-lughah al-arabiah bi ta'lim al-asasi. Kaherah: Dar al-Fikr

Wan Mohd Abrisam, Wan Ahmad (2002). Tinjauan sikap pelajar terhadap mata pelajaran Bahasa Arab Komunikasi di sekolah-sekolah daerah Machang, Kelantan. Bangi: Projek Penyelidikan Sarjana; Universiti Kebangsaan Malaysia.

Webb, S. (2005). Receptive and productive vocabulary learning: the effect of reading and writing on word knowledge. SSLA, 27, 33-52.

Zahriah, Husin (2011). Perbandingan pengajaran dan pembelajaran perbendaharaan kata Bahasa Arab mengikut kluster semantik dan kluster bebas. Tesis sarjana: Universiti Kebangsaan Malaysia.