

**Word Frequency Level and Lexical Coverage in the  
Reading Comprehension Texts  
of the Malaysian University English Test**

---

---

**Melisa Charles BENEDICT**

School of Languages, Civilisation and Philosophy  
Universiti Utara Malaysia

Email: [m\\_cb109@yahoo.com](mailto:m_cb109@yahoo.com)

**Ahmad Affendi SHABDIN**

School of Languages, Civilisation and Philosophy  
Universiti Utara Malaysia

Email: [affendi@uum.edu.my](mailto:affendi@uum.edu.my)

**Abstract**

The central foundation of knowledge is through grasping what is being read. Therefore, it is significant to recognise the difficulty level of a reading text for better and comprehensive understanding. The present study is an attempt to identify the word frequency level and lexical coverage of the Malaysian University English Test (MUET) reading texts which is deemed to be intricate for students. The anticipation of the Education Ministry of Malaysia is that students are capable and competent in comprehending academic texts which is required of them at the university. Nonetheless, the problem lies in students' lack of vocabulary size and knowledge which hampers their understanding of the reading texts. Thus, utilising 18 MUET reading texts from three past year sittings enabled the identification of the texts difficulty employing the quantitative approach. The data was generated using the Web Vocabulary Profiler (VP-Compleat Lexical Tutor) to obtain the frequency levels and percentages. The findings revealed that mastery of a minimum 6,000-word family level is needed to reach the lexical coverage of 95% and more than 8,000-word level to reach the 98% line. In terms of

the texts similarities and differences throughout the sittings, there were unnoticeable differences in the figures which indicates that the MUET texts demonstrated a consistency in its selection throughout all the three sittings. This study accomplishes an extensive thought towards the pedagogical implication and significant strategies for further intervention.

**Keywords:** *MUET, reading comprehension text, word frequency level, word family, lexical coverage*

## **Introduction**

The success of comprehending a reading text lies in the size and knowledge of words. According to Nuttall (1996), a good reader is able to rapidly identify words. This can only mean that a good reader possesses good vocabulary in terms of their breadth and depth vocabulary in order to recognise words to comprehend their meanings. Oyetunji (2011) stated that comprehension is vital in reading. Certain types of words found as significant in the writing can help readers to determine the definitions that cover the most authentic or relevant reasons for the reading. To improve learner's understanding of texts and their performance in the second language, the ways second language (L2) reading are taught is crucial.

Based on the foreword in the test specification of the MUET syllabus (2006, 2011 & 2015), MUET has aimed to quantify pre-university students' proficiency level in the English language for entrance into tertiary level. Moreover, it is acknowledged as a standardised proficiency test similar to IELTS and TOEFL and is also internationally recognised to reliably measure learner's English language capability as stated by Rethinasamy and Chuah (2011). It is the most significant English language tests targeted mainly for Malaysian pre-university students who are preparing for higher education (Othman & Nordin, 2013). It should be noted that all these proficiency tests quantify the capability of students to operate and understand English in college or university campuses. In sum, the MUET is a proficiency test that segregates and identifies the good and low proficiency students as a pre-requisite for university entry based on the courses chosen. Since MUET is a high-stake test, it needs to be addressed as well as taken seriously.

The MUET reading test description indicates that the text selection for the MUET reading comprehension is based on the level of complexity of content and knowledge as well as type of text. This has led to article selections from various electronic texts, newspapers, journals, academic texts and magazines from around the world which meet the level of complexity required in terms of its content and language. Based on the analysis reported by Hamzah (2013), officers from the Malaysian Examinations Council advised candidates sitting for the examination to attempt the reading of the passages in the order in which they appear in the test paper since the difficulty level of each passage increases in that order. Therefore, this research focusses on the word frequency level and lexical coverage of the MUET reading texts. This is for the reason that MUET reading comprehension being the component which holds the highest score of 120 marks (40%) upon 300 marks is definitely a challenge for students to score and knowing the level of complexity of the reading texts in MUET it is deemed necessary to aid students in expanding their vocabulary and most significantly text comprehension.

## **Literature Review**

### **Vocabulary in Second Language Learning**

Nation and Waring (1997) claimed that the number of words that an L2 learner requires rely on how the learner desires to use the language. However, to comprehend authentic second language texts, a wider vocabulary size is required with a minimum range of 3000 to 5000 words. However, to read an academic text, a learner needs to comprehend vocabulary size within a range of 8000–9000 word families (Schmitt, 2008). According to Nation (2010) in order to comprehend a variety of authentic material, a vocabulary of 8000-9000 word families is required. Therefore, Schmitt (2008) noted that L2 learners need to make every effort to expand their size of vocabulary if they desire to read a wide range of materials with less interference by unfamiliar vocabularies.

It was stated that a huge vocabulary is vital to function in English: 8000—9000 word families for reading (Schmitt, 2008). Thus, mastery of a certain vocabulary size is needed to understand

what is being read. According to Tseng & Schmitt (2008), the construct of vocabulary is quite complex. They assured that certain vocabulary sizes are required to perform certain things in language based on an extensive range of research conducted on vocabulary. The lexical requirements for English can be summarized as follows: 3000 word families to begin reading authentic texts, 5000-9000 word families to be independently reading authentic texts and 10,000 word families to allow most language use.

Researches on the relationship between the percentage of vocabulary known and the comprehension level of the same text had revealed that second language learners need to understand around 98% of the running words as a more equitable coverage of academic text for unassisted comprehension (Hu & Nation, 2000; Schmitt, Jiang & Grabe, 2011). Based on the research done on corpora of various genres, it was noted the value of 98% is parallel to 8,000 – 9000 word families (Nation, 2006). Therefore, these figures require a great deal of effort from second language learners for cautious and incidental vocabulary learning.

### **Word Frequency Levels**

The frequency of a word that occurs in a language is vital as words which occur more frequently can enhance one's vocabulary. According to Nation and Waring (1997), if learners know the words of English which occur frequently especially content words, they will be able to comprehend a great amount of running words in a written or spoken text. Possessing adequate knowledge of content words will lead to a good degree of comprehending a text. Moreover, learners can obtain the best return in their attempt to learn vocabulary if the frequency information is reasonably provided accordingly. It was mentioned that educators should take into consideration of the list of words which require attention or vice versa and the suitability of texts to be used in classroom as vocabulary frequency list plays an essential role in determining learning goals (Nation & Waring, 1997).

Based on Laufer's (1989) empirical research, students who were able to recognise 95% of the words in a text would be inclined to score 60% on a comprehension test for the text. Whereas, Nation's (2006) research indicated that students who were able to

know 98% of the words in a text had the tendency to score 70%. Furthermore, knowing 95% of the words in a text would be parallel to having vocabulary size of 5,000-word families for an average text in contrary to Nation's (2006) idea that 8,000-word families would be the proposed vocabulary size to gain the 95% figure. Both these notions depend on the type of text to a certain extent.

In two related studies (Aziez, 2011; Aziez, Furqanul & Aziez, Feisal, 2018) on junior and senior high school English National Examination (NE) texts in Indonesia, using the distribution of 1,000 to 20,000 vocabulary level, it was reported that the national examination texts of the junior and senior high school fell into the 4,000-word level, in which a level that is needed for a 95% understanding of the NE texts. This result showed that students needed a minimum of 4,000 vocabulary level at minimum in order to comprehend the NE reading texts.

Moreover, the impact and relationship between vocabulary knowledge and reading comprehension can also be seen in the study by Kameli and Baki (2013) on 220 second semester English as a Foreign Language (EFL) Iranian adult students at a private English Language Institution in Iran. The use of the Vocabulary Levels Test (VLT) and Reading Comprehension Test (International English Language Testing System or IELTS) revealed that the different levels of vocabulary had a positive relationship with the test scores in terms of vocabulary knowledge and reading comprehension. This finding supports the study by Harji, Balakrishnan, Bhar and Letchumanan (2015) on 120 first year Malaysian undergraduates at a private university in Malaysia whereby English was the medium of instruction. The researchers reported that almost none of the students managed to obtain more than 2,000 word-level which did not meet the required level of the University Word List (UWL). Hence, the findings depicted that the students' vocabulary knowledge was at a lower scale and inadequate to deal with reading texts and probably face difficulties to cope with studies at the university as well.

On the other hand, another research on vocabulary level and reading by Tan and Goh (2017) on 53 second year students pursuing their studies at a private university in Malaysia revealed that the vocabulary size of the students was only average with just

over 6000 word families using the Vocabulary Size Test and the IELTS Reading Test. This finding concurs with the study done by Lateh, Shamsudin and Raof (2018) that majority of the Malaysian university undergraduate students in their study demonstrated a weak possession of receptive vocabulary knowledge. It was reported that 93% of the undergraduate students failed to reach the mastery of the 5,000-word level including the academic word level. These findings are also in line with a study carried out by Ibrahim, Sarudin and Muhamad (2016) that only half (54.3%) of the pre-university students in a public university in Malaysia attained the mastery level of 5,000 word families based on the vocabulary levels test.

The results indicated that the vocabulary size possessed by these students was deficient to comprehend reading texts based on Nation's (2006) argument of 8,000 to 9,000-word level required for adequate comprehension to achieve the 98% figure and a minimum of 5,000-word level to reach the 95% line (Laufer, 1989; Van Zeeland & Schmitt, 2012). These researchers stated that the 10,000 word-family level was needed to be proficient in reading comprehension. In general, it was said that a reader should comprehend at least 19 of 20 words in average to reach the 95% lexical coverage for reading a text (Chujo & Oghigian, 2009).

Moreover, Chen (2011) asserted that there is a positive and significant correlation between vocabulary knowledge and reading comprehension in which learners who are of higher language proficiency were better in deducing the content of the reading texts compared to learners who are of lower language proficiency. Nevertheless, Nouri and Zerhouni's (2018) research reported a significant high correlation between vocabulary size, vocabulary depth and reading comprehension and that word frequency can effectively determine the difficulty of the reading material.

Chujo and Oghigian's (2009) study identified that the vocabulary size needed by learners in order to comprehend three different types of proficiency tests was a minimum vocabulary size of 3,000 word families to reach the 95% line on the Test of English for International Communication (TOEIC), 3,500 word families for the Test of English as a Foreign Language (TOEFL) and 4,500 word families for the Test in Practical English Proficiency (EIKEN). These

findings yielded an almost similar pattern and findings with other researches (Aziez, 2011; Aziez, Furqanul & Aziez, Feisal, 2018).

However, to date, little has been known and researched regarding the word frequency level and lexical coverage of the MUET reading texts. Thus far, only a few studies have been conducted regarding MUET reading text. Ong, Krishnan, Christopher Selvaraj and Renu (2015) had explored the skill of reading in relation to the MUET reading text based on its readability and text selection. Meanwhile, other recent research (Ong & Yuen, 2015, 2017 in press) conducted studies on the lexical bundles found in the MUET reading texts. Due to the lack of research in this area especially in the Malaysian context, the current study is deemed necessary to identify the MUET text difficulty for efficacious comprehension. Based on past literatures, it can be determined that the vocabulary size in which a learner possesses plays a crucial role in determining the success of comprehending a reading text. However, determining the vocabulary frequency levels in a reading text is deemed crucial as it speaks of text difficulty and vocabulary density or volume. Most of the MUET reading texts comprise sentences which are long and complex that contributes as a factor to the difficulty of the texts and the number of difficult texts were almost the same across 16 years since the implementation of the MUET (Ong, Krishnan, Christopher Selvaraj and Renu, 2015). It is therefore significant to identify MUET text difficulty through the vocabulary level and lexical coverage in order to aid students where it is reckoned necessary as far as reading is concerned.

According to Nation and Anthony (2013), vocabulary frequency levels can be divided into 3 broad frequency levels; low frequency, mid-frequency and high frequency. Taking into consideration of Schmitt and Schmitt (2012) word frequency level, Nation and Anthony (2013) reported three levels of vocabulary frequency levels as can be seen in Table 1.

**Table 1. High-frequency, mid-frequency, and low-frequency vocabulary**

Vocabulary level	Word family levels (and total)	Nature of the vocabulary
High-frequency	1st 1000-3rd 1000 (3,000)	Wide range, very high-frequency, essential, general purpose vocabulary
Mid-frequency	4th 1000-9th 1000 (6,000)	Wide range, moderate frequency, general purpose vocabulary
Low-frequency	10th 1000 on	Narrower range, low-frequency, some technical vocabulary unique to a particular discipline

Nation and Anthony (2013)

In Table 1, the high-frequency vocabulary consists of the 1<sup>st</sup> 3000-word family level followed by the mid-frequency vocabulary which consists of 4000 to 9000-word family level with a total of 6000-word families. The low-frequency vocabulary however was set from the 10,000-word family level onwards. Nation and Anthony (2013) reported that the cut-off point of the low frequency level was at 9000-word family since 9,000-word family presents 98% coverage of most texts. The three broad frequency vocabulary levels are vital to identify vocabulary levels and lexical coverage of reading texts.

### **The Corpora**

Corpus or corpora is a collection of texts in a large scale. It is a linguistic analysis based on the written or spoken texts. Some of the most popular corpora are British National Corpus (BNC), COBUILD/Birmingham Corpus, IBM /Lancaster Spoken English Corpus (Robin, 2009). The most reliable and updated corpus is the Corpus of Contemporary English (COCA) (Davies, 2010). The use of corpus is to supply information on lexical, syntax, semantic and pragmatic aspects. It aids grammarians, lexicographers and other researchers to obtain detailed description of a language (Robin, 2009). According to Xu (2014), corpora permits access to authentic data and show construction of grammar construction and patterns in terms of its word frequency. The patterns analysed can aid in the improvement of language material for teaching students. Moreover, corpora which are analysed via computer help linguists to inspect or retrieve information of a particular or selected text in terms of its lexis and word structure. Linguistic information such as frequency



count and concordance can be obtained through computer-assist (Bird, Klein & Loper, 2009).

*British National Corpus (BNC).* The British National Corpus (BNC) has running words of 100,000,000 in English consists of 10% from spoken sources and 90% from written sources including the Academic Word List (AWL) that occurs in the BNC lists. It was noted that the BNC covers variety of texts and corpora better than other lists. According to Nation (2004), an additional of 444 word families were found in the BNC 3000 list more than the GSL and AWL which proofs that the BNC has a better coverage. Nation (2004) mentioned that the first 2000-word family encompasses words found in the AWL while the GSL is mostly non-academic words. However, Davies (2010) stated that the BNC has not been updated since it was developed and completed in 1993. The corpus covers texts from the 1970's up to the early 1990's, nevertheless no new texts has been added to this corpus since then and nor will it be added up in future (Davies, 2010).

*Corpus of Contemporary American English (COCA).* Corpus of Contemporary American English (COCA) has been claimed by Davies (2010) and Xu (2014) to be the first most reliable and balanced corpus of English as it has been divided almost equally (20% for each genre) for spoken, popular magazines, fiction, academic journals and newspapers. Besides, the data of COCA is made available for changes which are ongoing in the English Language which are inaccessible from other sources. It is the only corpora which is continuously updated unlike Bank of English (BoE) and Oxford English Corpus (OEC) which stopped their updates in 2005 and 2006 respectively.

Besides, COCA has an addition of 20 million words added every year since 1990 in which the data can be compared between different years as well as time period. COCA updates itself to changes and what is happening in the actual world. COCA provides users with much convenience of searching part of speech, synonyms, collocates, lemma etc. and provides ample lexis and grammar patterns for the required word frequency and its usage (Davies, 2010; Xu, 2014). With COCA, teachers can analyse the way

the language is changing which is not possible with other resources (Davies, 2010).

In brief, the related reviews on corpus were necessary as both the corpora discussed were used in identifying the word frequency level for MUET reading passages based on the BNC and COCA corpora.

### **This Study**

This study seeks to examine the word frequency level and lexical coverage of the MUET reading text depicted by the word frequency levels from 1000-word level (K-1) to 25,000- word level (K-25). An analysis was conducted to determine the level of difficulty of the MUET reading texts for the past few sittings. It is necessary to note that the focus of this current study is not to probe into the vocabulary size of test-takers or test items of the reading texts but to solely explore the word frequency level and lexical coverage of the several genres utilised for the MUET reading texts that are sourced from articles taken from newspapers, magazines, journals as well as academic texts and electronic texts in preparing students to face such texts. Therefore, the present study looks into the aspect of difficulty of the MUET reading texts in terms of its word frequency level and lexical coverage.

### **Research Questions**

The research questions for this study are as follows:

- a) To what extent does the high, mid- and low frequency word levels are covered in the Malaysian University English Test reading texts (distribution between the K-1 to K-25 word levels)?
- b) How much vocabulary does a learner need to comprehend MUET reading texts based on the word frequency level and to what extent is the lexical coverage of the three reading comprehension sittings?
- c) What are the similarities and differences of the 18 reading texts across all the three MUET sittings?

## **Methodology**

Prior to answering the research questions, the progress of the study went through a few key steps. First and foremost, the 18 MUET reading texts from 3 different sittings (July 2015, November 2015 and March 2016) were selected to identify the vocabulary level of each text. The 18 texts from the three separate sittings were chosen and represent three separate sittings each year as most of the MUET reading texts were deemed to be of an equal level of complexity across the years based on past research (Ong, Krishnan, Christopher Selvaraj and Renu, 2015). Next, the MUET reading comprehension texts in the form of printed pages were scanned into pdf .jpeg format and converted to word document files. Irrelevant and unnecessary graphics and questions were removed to only retain the text data for data generation. This method was adapted from Ong and Yuen (2015) research procedure. Subsequently, the word frequency levels of the reading texts were generated using the Web Vocabulary Profiler program (VP-Compleat Lexical Tutor) developed by Cobb (2004). Lastly, the percentage of the words families and tokens based on the word frequency levels were analysed for lexical coverage.

### **The Instrument**

In this study, the Vocabulary Profiler (VP-Compleat Lexical Tutor) and MUET reading texts from three different sittings; July (2015), November (2015) and March (2016) were utilised. In general, Vocabulary Profilers generate the word frequencies in a corpus by breaking the texts down. The Web Vocabulary Profiler program (VP-Compleat Lexical Tutor) serves as an instrument that checks whether a piece of text comprises words from vocabulary list as well as generates the word frequency levels based on the word families. It contains the BNC and COCA corpora that was used in this study. It specifically checks and generate the level of vocabulary and difficulty of a reading text and later indicates the number of words the texts contain from the frequency bands. It provides the word level frequency up to 25,000-word level and determines the difficulty level of each text.

### **Analysis of the MUET Reading Texts**

The next step involves the analysis of the percentage of the MUET reading texts with special emphasis on the word frequency level based on the frequency bands and cumulative token percentage for comprehension coverage using the Web Vocabulary Profiler (VP-Compleat Lexical Tutor). The aim was to identify the level of the frequency band that a learner must attain to achieve the desired comprehension coverage. The analyses were conducted and results were generated based on the British National Corpus (BNC) and Corpus of Contemporary American English (COCA) which were created in the profiler. The analyses started with the selection of each reading comprehension text from all the three sittings (18 texts) and each text was later generated separately in the Vocabulary Profiler to obtain the results for the word frequency level and cumulative tokens. The texts are of analytical, descriptive, persuasive, argumentative and narrative types of texts which covers science and arts-based discipline either in the fields of life sciences, earth sciences, performing arts, environmental studies, geography, international trade, agriculture, literature, physics, chemistry, architecture and design, entrepreneurship, history and gender studies (Ong & Yuen, 2015). Consequently, the use of the British National Corpus (BNC) and Corpus of Contemporary American English (COCA) in the vocabulary profiler was to look into the lists of word frequencies that serve a purpose in this research as the MUET texts are mostly taken from British and American sources.

### **Results and Discussions**

#### **The extent of high, mid- and low frequency words used in the MUET reading texts based on the word frequency distribution between the K-1 to K-25 word levels**

This section examines the vocabulary level for each MUET reading comprehension sitting based on the high, mid and low word frequency levels as can be seen in Table 2. Based on the analysis conducted on *Reading Texts One to Six* for July 2015, it was found that the word frequency level of these texts reached up to K-17 (17,000) word family level. For these texts, the high-frequency

vocabulary level for the 1<sup>st</sup> 3000 (K-3) word family based on all six texts reached up to 91.09% for each text. As for the mid-frequency level (4000 to 9000- word family level), the vocabulary reached up to only 12.72% and the low frequency level which consists of the 10,000- word level onwards was only up to 0.68%.

**Table 2. Summary of High, Mid and Low Vocabulary Frequency Level**

Vocabulary Level	Word family level	July 2015	November 2015	March 2016
High-frequency	K-1 to K-3 (1000-3000)	up to 91.09%	up to 93.92%	up to 98.36%
Mid-frequency	K-4 to K-9 (4000-9000)	up to 12.72%	up to 10.6%	up to 9.9%
Low-frequency	K-10 and above (<10,000)	up to 0.68%	up to 1.5%	up to 2.24%

K= '000

However, for the November 2015 sitting, the word frequency level of the reading texts reached up to K-24 (24,000) word family level. It can be seen that the high-frequency vocabulary level reached more than 93.92% which did not make much difference with the July 2015 sitting. For the mid-frequency level, the vocabulary reached up to 10.6% and for the low frequency level, it reached up to only 1.5% which also denoted not much of a difference. In comparison to the two previous sittings, the reading texts for March 2016 sitting showed no difference in its word frequency level as the texts for this sitting also reached up to K-24 (see Table 3). However, looking at the high frequency level vocabulary, the percentage touched 98.36% which was considerably high whereas the mid- frequency and low frequency levels made up to 9.9% and 2.24% respectively.

Nevertheless, looking at the extent of the three vocabulary levels used in the MUET reading texts based on the vocabulary levels outlined by Nation and Anthony (2013), it can be summed up that the difficulty level of the MUET reading texts remained the same throughout the three sittings and most prominently the passages chosen were consistent in terms of their word frequency levels based on the three levels divided. It can be noted that the vocabulary used in all the reading texts for MUET fell in the 1<sup>st</sup> 3000-word family level which denoted that more than 90% of the word families were utilised within this range. The findings indicated

that as the word family level increases, the lower the percentage of the vocabulary level becomes. As far as the 18 reading texts were concerned, all three sittings showed that the word families which fell in the K-3 (3000) word level for each passage were between the range of 91% to 98%. Nevertheless, up to 12.72% and 2.24 % of the words fell in the range of K-4 (4000) to K-9 (9,000) and K-10 to K-25 (10,000- 25,000) word levels respectively. Therefore, it can be concluded that most of the words fell in the high frequency band for all the three sittings.

Although this result demonstrated that around 90% of the words fell within the 3000-word family, the lexical coverage of minimum 95% in order to comprehend the MUET reading texts was not achieved at this level. The lexical coverage within the 3,000 word-level was only found to reach mostly below the 95% line (80%-90%) for 15 texts out of the 18 reading texts (refer Tables 4 and 5). Chujo and Oghigian (2009) study found that a learner needs a minimum vocabulary threshold of 3,000 word families to reach the 95% line on TOEIC and 3,500 word families for TOEFL. Nonetheless, those findings did not yield the same result with the current study since the 95% lexical coverage was not covered at the 3,000- word level for MUET albeit they are of the similar standardised proficiency tests. Therefore, that a higher vocabulary level is needed to attain the minimum 95% line for the MUET reading texts as compared to the other two tests. This could feasibly mean that the MUET reading texts are considered tougher to some extent in comparison to TOEIC, IELTS and TOEFL as the 95% lexical coverage can only be attained at a higher word band of more than 3,000-word level. Therefore, an in-depth analysis of the word frequency levels and lexical coverage that should be attained based on each MUET reading text for text comprehension will be presented in the next section.

### **Vocabulary size and percentage of the lexical coverage needed in comprehending the MUET reading texts.**

This section analyses the lexical coverage and the possible vocabulary threshold required by students to comprehend the MUET reading text. A detailed analysis on the word frequency level

and lexical coverage for each individual reading comprehension texts were computed as compared to the previous section which dwelled into a holistic word frequency level of all the 6 reading comprehension texts for each sitting. In this study, vocabulary size refers to the number of words in the word family that pre-university students need to possess and this is indicated by the word frequency level. The percentage of lexical coverage (indicated in the form of cumulative tokens) refers to the proportion of words that must be known by students to comprehend a text. It is interesting and vital in this study to identify and analyse the vocabulary level of reading texts and the lexical coverage that learners need to know for successful comprehension. To answer this research question, each reading texts were generated based on the BNC-COCA K-25 (25,000) word frequency level and the findings were summarised in Table 3. These findings are based on the lexical coverage cutting point of 95% and 98% by Van Zeeland & Schmitt (2012) indicating that for ‘adequate’ comprehension 98% is considered as sufficient and 95% as minimum lexical coverage. Moreover, Schmitt, Cobb, Horst and Schmitt (2017) suggested that to attain the 98% of lexical coverage, learners are required to know around 8,000 to 9,000-word families including proper nouns.

**Table 3. Summary of the Cutting Point of the Word Frequency Level and Lexical Coverage for MUET Reading Text.**

Reading Text	July 2015		November 2015		March 2016	
	WFL	CT	WFL	CT	WFL	CT
1	K-13	89.97	K-11	93.94	K-3	95.42
2	K-16	93.07	K-24	92.25	K-12	94.60
3	K-12	95.07	K-8	95.23	K-19	90.78
4	K-7	96.18	K-12	95.19	K-15	95.42
5	K-14	95.09	K-3	95.50	K-3	95.15
6	K-7	95.09	K-6	95.30	K-7	95.28

WFL= Word Frequency Level, CT= Cumulative Token represents lexical coverage

VP-Compleat Lexical Tutor indicated that if the 95% line is attained at only the 1,000-word family level, then the text is deemed as fairly basic text for the students. However, it was mentioned that if the 95% figure is only attained at the starting of 5,000 or 6,000-word level, it can be deduced that the text has difficult vocabularies in which is not suitable for low proficiency students.

Therefore, based on the percentage of lexical coverage set by the researchers, the findings showed that for the July 2015 MUET reading comprehension texts, the cutting point for the word frequency level and the intended 95% lexical coverage at minimum were found in Reading Texts 3 to 6. It can be seen for Reading Texts 3 and 5, students need to acquire at least 12,000 and 14,000-word families respectively at minimum in order to successfully comprehend both the reading texts. Hence, these two texts consume complex vocabularies and deemed to be somewhat difficult to comprehend if they fail to possess the minimum 12,000-word level. However, for Reading Texts 4 and 6, students need to possess at least 7,000- word families. Nonetheless, looking at the word-level set for 95% comprehension coverage at minimum, these texts are considered as difficult.

For the November 2015 paper, it was also found that Reading Texts 3 to 6 reached the lexical coverage of 95 %. Looking at the word families that students need to possess, the findings depicted that students need to possess a vocabulary level of up to 12,000-word family for Text 4 and a minimum of 3,000-word family for Text 5. However, for the other reading texts, students need to acquire between 6,000 to 8,000- word families. This indicates that Text 5 seemed to be easier and can be signified as a fairly average text to comprehend since the VP-Compleat Lexical Tutor indicated that if the 95% figure is attained only at the 1,000-word family level, the text is fairly basic. As for the other texts, the complexity of the vocabularies can be clearly seen.

In comparison to the March 2016 paper, the findings portrayed that the 95% line was depicted in Reading Texts 4 to 6. Reading Text 4 reached the 95% lexical coverage at the 15,000-word family which means students attempting this text need to have a vocabulary size of such to comprehend the text. Therefore, a second language learner who attempts this text would need to have a wide range of vocabulary to comprehend the text at a minimum level. Conversely, for Text 5, the 95% lexical coverage was attained at the 3,000-word level which can be considered as an average text to comprehend compared to Text 6 coverage in which the 95% line started only from the 7,000-word level thus considered as a difficult text.



It is also significant to note that most of the texts are indeed difficult and challenging for second language learners. Nevertheless, as the reading texts move from Text 1 to Text 6, the vocabulary level of the texts gets tougher as can be noted in Table 3. This finding is in line with Hamzah (2013) who mentioned that the MUET reading texts should be attempted in the order in which they appear in the test paper since the level of difficulty of each text increases in that order.

In sum, the findings for this research question revealed that the texts are undeniably testing the students' proficiency level. If they do not reach the minimum required vocabulary level for text comprehension, chances are students might not be able to comprehend the MUET reading text successfully for better reading performance. It can be clearly seen that students need to attain at least 6,000-word family level for text comprehension at minimum as out of the 18 Texts, 15 Texts consumed of vocabularies starting at the level of 6,000-word family for 95% lexical coverage. Based on Laufer's (1989) empirical research in which if students have reached the 95% line, they would score 60% of the reading comprehension questions. This means, with a higher lexical coverage, a higher score might be obtained. Hence, comparing with the lexical coverage of the MUET reading texts and the word frequency level that students need to attain, they must make the necessary effort to boost their vocabulary level up to 6,000-word family level at minimum in order to achieve at least 60% correct for the reading comprehension text. However, if they desire to achieve more than 60% correct, they need to achieve a higher vocabulary level of more than 8,000 word-level as stated by Nation (2006); Van Zeeland & Schmitt (2012); Schmitt, Cobb, Horst and Schmitt (2017) to reach the 98% lexical coverage. This means that pre-university students sitting for the MUET should acquire more than 8,000 word families to reach the 98% coverage and a minimum of 6,000-word level to reach the 95% line in order to be a proficient reader.

Past studies have revealed that Malaysian undergraduate students demonstrated a weak possession of receptive vocabulary knowledge and failed to even reach the 5,000 vocabulary level based on vocabulary tests conducted (Ibrahim, Sarudin & Muhamad, 2016; Lateh, Shamsudin & Abdul Raof, 2018) and with such

vocabulary level at hand it would require students to work on quite a number of new words to comprehend the texts. A question arises whether pre-university and even university students are able to comprehend the MUET reading texts of such level. This question can be related to the findings by Tan and Goh (2017) that the vocabulary size of the students at a private university in Malaysia were average with just over 6000 word families. The results indicated that the vocabulary size possessed by these students was deficient to comprehend reading. Besides, Harji, Balakrishnan, Bhar and Letchumanan's (2015) study stated that almost none of the undergraduate students in their study managed to obtain more than 2,000 word-level which did not meet the required level of the University Word List (UWL).

Considering a minimum of 6,000 vocabulary threshold to comprehend the MUET texts, it is deemed quite difficult for second language pre-university students to comprehend those texts since students even at tertiary level have not reached at least a minimum vocabulary possession at 5,000-word level based on findings of past studies. However, observing the minimum 5,000 vocabulary level needed for text comprehension especially for tertiary education, the need for students to acquire this level of vocabulary size at minimum to comprehend books, journal articles, periodicals and other academic materials at university is vital. Hence, the level of MUET text difficulty can be seen in this study and found to be suitable for a national standardised examination to equip pre-university students for academic reading at tertiary level. Additionally, the capability of these students in mastering their English language skills to perform efficiently in their academic quest at higher educational institutions is the aim of the MUET syllabus. Apart from that, it is also the expectation of the Malaysian Ministry of Education that students should demonstrate the ability to further progress in their academic vocabulary, able to apply newly acquired vocabulary and to be autonomous readers as outlined in the Common European Framework of Reference for Languages (CEFR).

**Similarities and differences across all the 18 Reading Texts in all three MUET sittings.**

From the data generated using the VP-Compleat Lexical Tutor, a similar pattern could be identified in Tables 4, 5 and 6. The overall pattern of the word level frequency based on the BNC-COCA word lists fell in the 1<sup>st</sup>1000-word family. A huge gap existed between the first 1000 and the 2<sup>nd</sup>1000- word family level. The percentage of the 1<sup>st</sup>1000-word-family depicted more than 46.51% and reached up to 81.96% throughout all the 18 texts. However, the 2<sup>nd</sup> 1000-word family figures were between 12.74% and 23.64%. The other word levels showed only a slight margin between each frequency levels. The July 2015 texts reached up to the level of 17,000- word family whereas the November 2015 and March 2016 texts reached up to 24,000-word family level respectively. However, in terms of its differences, not much could be noticed from the figures generated as most of the figures were almost similar after the 2,000-word family level. Therefore, the MUET texts showed a consistency in their selection throughout all the three sittings.

**Table 4. Analysis of Word Frequency Level and Lexical Coverage of the MUET Reading Texts based on the BNC-COCA for July 2015**

Word Frequency level	Percentage of Word Families and Cumulative Token for Each Word Frequency Level											
	RP 1		RP 2		RP 3		RP 4		RP 5		RP 6	
	WF	CT	WF	CT	WF	CT	WF	CT	WF	CT	WF	CT
	%	%	%	%	%	%	%	%	%	%	%	%
K-1 Words	65.61	72.56	63.36	68.32	61.48	73.51	63.86	75.48	59.35	73.39	56.70	68.73
K-2 Words	12.74	80.74	13.79	77.82	16.80	83.57	17.27	85.99	18.69	83.33	16.84	79.84
K-3 Words	12.74	86.02	13.79	86.33	11.07	89.73	9.24	91.09	10.98	89.21	13.06	88.37
K-4 Words	3.18	87.34	2.16	87.72	2.46	90.96	3.61	93.32	2.97	90.75	4.81	90.57
K-5 Words	1.91	88.13	3.45	91.48	2.46	92.40	1.20	93.80	2.37	91.87	3.78	93.15
K-6 Words	1.27	88.66	-	-	1.64	93.22	2.01	94.91	0.89	92.43	2.41	94.83
K-7 Words	-	-	1.29	92.07	1.64	94.04	1.20	96.18	0.30	92.57	0.69	95.09
K-8 Words	0.64	88.92	-	-	-	-	0.80	96.50	1.48	93.41	-	-
K-9 Words	0.64	89.18	-	-	0.82	94.45	0.40	96.66	1.78	94.39	1.03	95.87
K-10 Words	-	-	0.86	92.47	-	-	-	-	0.59	94.67	0.34	96.00
K-11 Words	-	-	-	-	0.82	94.86	-	-	-	-	-	-
K-12 Words	0.64	89.71	0.43	92.67	0.41	95.07	-	-	-	-	-	-
K-13 Words	0.64	89.97	0.43	92.87	0.41	95.28	-	-	-	-	-	-
K-14 Words	-	-	-	-	-	-	-	-	0.59	95.09	-	-
K-15 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-16 Words	-	-	0.43	93.07	-	-	-	-	-	-	0.34	96.13
K-17 Words	-	-	-	-	-	-	0.40	96.82	-	-	-	-
K-18 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-19 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-20 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-21 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-22 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-23 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-24 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-25 Words	-	-	-	-	-	-	-	-	-	-	-	-

K= '000, RP= Reading Passage, WF= Word Family, CT= Cumulative Token

**Table 5. Analysis of Word Frequency Level and Lexical Coverage of the MUET Reading Texts based on the BNC-COCA for November 2015**

Word Frequency level	Percentage of Word Families and Cumulative Token for Each Word Frequency Level											
	RP 1		RP 2		RP 3		RP 4		RP 5		RP 6	
	WF	CT	WF	CT	WF	CT	WF	CT	WF	CT	WF	CT
	%	%	%	%	%	%	%	%	%	%	%	%
K-1 Words	59.69	67.67	56.85	63.15	61.09	71.55	65.78	80.93	46.51	58.09	57.41	73.22
K-2 Words	16.84	80.46	18.78	75.22	16.73	81.91	17.49	88.75	23.64	80.10	16.09	82.10
K-3 Words	13.78	88.83	14.21	83.19	8.73	87.01	10.65	92.87	22.09	95.50	15.14	89.46
K-4 Words	2.55	89.99	2.03	84.91	2.91	89.64	2.66	93.83	2.71	96.12	5.36	93.65
K-5 Words	4.59	92.78	3.05	86.42	4.36	92.93	1.14	94.24	3.10	97.41	2.21	94.54
K-6 Words	0.51	93.01	1.52	87.07	1.45	93.75	0.38	94.38	0.39	97.57	1.58	95.30
K-7 Words	0.51	93.24	-	-	2.18	94.74	0.76	94.65	0.39	97.89	0.32	95.43
K-8 Words	0.51	93.47	0.51	87.29	1.09	95.23			0.78	98.21	0.2	95.68
K-9 Words	-		1.52	88.37	0.36	95.39			-		0.95	96.06
K-10 Words	-		0.51	91.39	0.36	95.55	0.76	94.92	-		-	
K-11 Words	1.02	93.94	-		0.36	95.71			0.39	98.37	-	
K-12 Words	-		-		0.36	95.87	0.38	95.19	-		-	
K-13 Words	-		-		-				-		0.63	96.31
K-14 Words	-		-		-				-		-	
K-15 Words	-		-		-				-		-	
K-16 Words	-		0.51	91.82	-				-		-	
K-17 Words	-		-		-				-		-	
K-18 Words	-		-		-				-		-	
K-19 Words	-		-		-				-		-	
K-20 Words	-		-		-				-		-	
K-21 Words	-		-		-				-		-	
K-22 Words	-		-		-				-		-	
K-23 Words	-		-		-				-		-	
K-24 Words	-		0.51	92.25	-				-		-	
K-25 Words	-		-		-				-		-	

K= '000, RP= Reading Passage, WF= Word Family, CT= Cumulative Token

**Table 6. Analysis of Word Frequency Level and Lexical Coverage of the MUET Reading Texts based on the BNC-COCA for March 2016**

Word Frequency level	Percentage of Word Families and Cumulative Token for Each Word Frequency Level											
	RP 1		RP 2		RP 3		RP 4		RP 5		RP 6	
	WF	CT	WF	CT	WF	CT	WF	CT	WF	CT	WF	CT
	%	%	%	%	%	%	%	%	%	%	%	%
K-1 Words	74.59	81.96	68.06	72.14	56.04	66.56	59.49	71.24	68.66	81.67	59.11	72.35
K-2 Words	16.39	91.44	14.14	83.16	18.68	77.07	16.08	80.51	15.30	90.16	17.89	83.62
K-3 Words	7.38	95.42	9.42	88.15	15.75	85.83	12.22	86.69	8.96	95.15	12.78	91.09
K-4 Words	0.82	95.73	3.66	89.81	5.13	88.70	3.22	88.17	2.61	96.23	3.83	92.92
K-5 Words	-	-	0.52	90.02	1.10	89.18	3.54	90.32	1.49	96.77	1.60	94.23
K-6 Words	-	-	0.52	91.06	1.83	89.98	0.32	90.45	0.37	97.04	1.28	94.89
K-7 Words	-	-	-	-	-	-	0.96	93.54	0.75	97.31	0.96	95.28
K-8 Words	-	-	2.09	92.93	-	-	0.64	93.81	0.37	97.44	0.32	95.41
K-9 Words	-	-	-	-	-	-	1.29	94.62	0.37	97.57	0.64	95.67
K-10 Words	-	-	-	-	0.73	90.30	-	-	0.37	97.70	0.32	95.80
K-11 Words	-	-	0.52	93.14	-	-	0.32	94.75	-	-	-	-
K-12 Words	-	-	1.05	94.60	0.37	90.62	-	-	0.37	97.83	0.62	96.19
K-13 Words	0.82	96.04	-	-	-	-	0.32	94.88	-	-	0.32	96.32
K-14 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-15 Words	--	--	-	-	-	-	0.32	95.42	-	-	-	-
K-16 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-17 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-18 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-19 Words	-	-	-	-	0.37	90.78	0.32	95.55	-	-	0.32	96.45
K-20 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-21 Words	-	-	-	-	-	-	0.64	95.82	0.37	97.96	-	-
K-22 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-23 Words	-	-	-	-	-	-	-	-	-	-	-	-
K-24 Words	-	-	-	-	-	-	0.32	95.95	-	-	-	-
K-25 Words	-	-	-	-	-	-	-	-	-	-	-	-

K= '000, RP= Reading Passage, WF= Word Family, CT= Cumulative Token

## **Conclusion and Pedagogical Implications**

The proportion of words known in a text is vital for adequate comprehension and for successful attempt of reading comprehension questions. As for pre-university students, they are required to comprehend texts of higher level of complexity from sources they have not been exposed to in MUET as preparation for their tertiary education. According to Mohd. Zin and Rafik-Galea (2010), substantial amount of time is spent by students to read academic materials and books at university. Thus, pre-university students must be prepared to read and understand materials such as from journal articles and academic books published in English to complete their assignments, projects and examinations at university. The maturity in understanding reading texts presented to them in a way measures their academic ability (Ong & Yuen, 2015).

The figures obtained from this study gave an insight into the MUET text difficulty and the gravity of the vocabulary load that students need to possess and sufficient vocabulary size can be clearly seen as much needed and vital for successful reading. The anticipation rests on students to expand their vocabulary knowledge at this level as well as the burden lies on educators to prepare them for this high-stake test as this test determines their entry into the university as well as for their desired courses. Therefore, the present findings are hoped to shed light on the importance of vocabulary size and knowledge in relation to reading comprehension. In revealing the lexical coverage and the vocabulary level students need to attain for successful reading through this study, it is significant for educators to take note of the challenges they might need to face in teaching reading at this level. Teachers who wish to teach reading comprehension successfully may consider the various vocabulary learning strategies and other reading strategies to aid students to better comprehend challenging reading texts. Choosing the appropriate and various strategies as well as teaching materials selections to aid reading comprehension through vocabulary can benefit the test takers and other learners of the same in unexpected ways.

## About the Authors

Melisa Charles Benedict is a lecturer at a Matriculation College and teaching pre-university students. She is specifically interested in vocabulary and reading in the second language. She is also a post-graduate student at the School of Languages, Civilisation and Philosophy at Universiti Utara Malaysia. Currently, she is pursuing her doctorate degree (Ph.D) at the university, specialising in Applied Linguistics. She can be reached at [m\\_cb109@yahoo.com](mailto:m_cb109@yahoo.com).

Ahmad Affendi Shabdin is an Associate Professor at Universiti Utara Malaysia from the School of Languages, Civilisation and Philosophy. He is a Ph.D holder in the area of Second Language Vocabulary Acquisition and Language Assessment. He teaches post-graduate programs at the university in Semantics and Language Assessment. He can be reached at [affendi@uum.edu.my](mailto:affendi@uum.edu.my).

## References

- Aziez, Furqanul (2011). Examining the vocabulary levels of Indonesia's English National Examination Texts. *Asian EFL Journal. Professional Teaching Articles*, 51, 16–29. Retrieved from <http://asian-efl-journal.com/PTA/April-2011-Aziez.pdf>
- Aziez, Furqanul & Aziez, Feisal (2018). The vocabulary input of Indonesia's English Textbooks and National Examination Texts for junior and senior high schools. *TESOL International Journal*, 13(3), 66–77.
- Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with python. Retrieved from <https://pdfs.semanticscholar.org/7a65/f23d990231d461418067c808b09d84c19b2c.pdf>
- Chen, K. Y. (2011). The impact of EFL students' vocabulary breadth of knowledge on literal reading comprehension. *Asian EFL Journal*, 51, 30–40.
- Chujo, K., & Oghigian, K. (2009). How many words do you need to know to frequency vocabulary. *The Journal of Asia TEFL*,



6(2), 121–148.

- Cobb, T. (2004). Web Vocab Profiler (Version 2.1) [Computer software]. Retrieved from <https://www.lex tutor.ca/vp/comp/>
- Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, 25(4), 447–464. <https://doi.org/10.1093/lc/fqq018>
- Hamzah, H. (2013, March 27). Working hard to improve English proficiency. *Bernama*. Retrieved from [http://education.bernama.com/index.php?sid=exclusive\\_content&id=937614](http://education.bernama.com/index.php?sid=exclusive_content&id=937614)
- Harji, M. B., Balakrishnan, K., Bhar, S. K., & Letchumanan, K. (2015). Vocabulary levels and size of Malaysian undergraduates. *English Language Teaching*, 8 (9), 119–130. <http://dx.doi.org/10.5539/elt.v8n9p119>
- Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 23(1), 403–430.
- Ibrahim, E. H. E., Sarudin, I., & Muhamad, A. J. (2016). The Relationship between Vocabulary Size and Reading Comprehension of ESL Learners. *English Language Teaching*, 9(2), 116. <https://doi.org/10.5539/elt.v9n2p116>
- Kameli, S., & Baki, R. (2013). The Impact of Vocabulary Knowledge Level on EFL Reading Comprehension. *International Journal of Applied Linguistics & English Literature*, 2(1), 85–89. <https://doi.org/10.7575/ijalel.v.2n.1p.85>
- Lateh, N. H. M., Shamsudin, S., & Abdul Raof, A. H. (2018). Receptive Vocabulary Levels of Malaysian University Students. *LSP International Journal*, 5(1), 105–113.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Clevedon: Multilingual Matters
- Mohd. Zin, Z., & Rafik-Galea, S. (2010). Anxiety and Academic Reading Performance among Malay ESL Learners. *Journal*

*of Pan-Pacific Association of Applied Linguistics*, 14(2), 41–58.

- Nation, P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). Retrieved from [https://www.lextutor.ca/cv/bogaards\\_laufer\\_2004.pdf](https://www.lextutor.ca/cv/bogaards_laufer_2004.pdf)
- 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>
- Nation, I.S.P. & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N.Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge, UK: Cambridge University Press.
- Nation, I. S. P., & Anthony, L. (2013). Mid-frequency readers. *Journal of Extensive Reading*, 1, 5–16.
- Nouri, N., & Zerhouni, B. (2018). Lexical frequency effect on reading comprehension and recall. *Arab World English Journal*, 9(2), 234–250. <https://dx.doi.org/10.24093/awej/vol9no2.1>
- Nuttall, C. (1996). *Teaching reading skills in a foreign language*. London: Heinemann.
- Ong, C. S. B., Krishnan, V., Christopher Selvaraj, J. J., & Renu, K. (2015). Readability of Muet reading comprehension passages. In *SOLLs.INTEC 2015 Proceedings* (pp. 92–101).
- Ong, C.S. B. & Yuen, C.K. (2015). Functional types of lexical bundles in reading texts of Malaysian University English test: A corpus study. *GEMA Online® Journal of Language Studies*, 15(1).
- Ong, C. S. B., & Yuen, C.K. (2017). Comparing structural and functional lexical bundles in MUET reading test. *Pertanika Journal of Social Sciences and Humanities*, 25( 1), 133–148.
- Othman, J., & Nordin, A. B. (2013). MUET as a predictor of academic achievement in ESL teacher education. *GEMA Online Journal of Language Studies*, 13(1), 99–111
- Oyetunji, C. O. (2011). *The effects of reading strategy instruction*

- on L2 teacher trainees' performance. University of South Africa. (Unpublished Master's Thesis) Retrieved from <https://core.ac.uk/download/pdf/43170904.pdf>
- Rethinasamy, S., & Chuah, K. M. (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.
- Robin. (2009). What is Corpus? *Natural Language Processing*, 1. Retrieved from <http://language.worldofcomputing.net/linguistics/introduction/what-is-corpus.html>
- Schmitt, N. (2008). *Review article: Instructed second language vocabulary learning. Language Teaching Research* 12 (3), 329-363. <https://doi.org/10.1177/1362168808089921>
- Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching. Advance online publication.* <http://doi.org/10.1017/S0261444812000018>
- Schmitt, N., Jiang, X., & Grabe, W. P. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26-43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 50(2), 212–226. <https://doi.org/10.1017/s0261444815000075>
- Tan, A.W.L & Goh, L. H. (2017). Relationship between vocabulary size and reading comprehension levels of Malaysian tertiary students. *International Journal of English Language & Translation Studies*. 5(4), 149-155.
- Tseng, W.-T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning*, 58(2), 357–400. Retrieved from <http://dx.doi.org/10.1111/j.1467-9922.2008.00444.x>
- Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479. <https://doi.org/10.1093/applin/ams074>

Xu, Y. (2014). On the application of Corpus of Contemporary American English in vocabulary instruction. *International Education Studies*, 7(8), 68–73.  
<https://doi.org/10.5539/ies.v7n8p68>