International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 14, Issue 1, 2020

Vocabulary Threshold for the Comprehension of Malaysian Secondary Engineering Texts as Compared to the Non-Engineering Genres

Y.J. Ng[a*], S.T. Chong[b], S. Thiruchelvam[c], M.F. Chow[d], J. Karthikeyan[e],
[a,b,c,d]Universiti Tenaga Nasional, Selangor, 43000, Malaysia, [e]Vellore Institute of Technology, Tamil Nadu, 632014 India, Email: [a*]yujin@uniten.edu.my

Learning technical content like engineering subjects using English could be overwhelming to secondary school learners whose English is a second or foreign language. The success of context understanding or text comprehension is often related closely to the vocabulary size of the learners and the extent of lexical coverage of the intended materials in terms of word families. The objective of the current study is to assess the required vocabulary size for the comprehension of engineering texts measured in the percentage of lexical coverage as compared to those of other disciplines at the secondary level. The study shows that all the engineering materials needed at least 10,000 word families' threshold to reach 95% coverage which is much higher as compared to the others. It is implied that the engineering texts require a larger vocabulary size for students to cope with understanding the texts. Thus, teachers who need to teach English for Engineering Purposes (EEP) should prepare pedagogically to cater to students' vocabulary needs.

Key words: *Vocabulary threshold, vocabulary size, engineering texts, engineering corpus, lexical coverage.*

Introduction

Word knowledge can be defined as a range of connected 'sub-knowledge' like grammatical and morphological knowledge, as well as an understanding of word meanings (Nation, 2011). Vocabulary or lexis relates to the semantics of a target language, in other words, it refers to the knowledge of words and the meanings, which forms the essence of a language learning experience. Hsu (2014) believes that word or lexical knowledge involves progressive learning of levels of knowledge from the most familiarised words and is able to use them in free production or what we call productive vocabulary. One can only function well in a language if he or she possesses the required vocabulary and thus, vocabulary is often regarded to be the key element in the success of second language learning (Coxhead, 2000). Nation (2006) stated that the concept of vocabulary is divided into two dimensions which are the knowledge of the words and word skills. Word knowledge refers to the understanding of word-form including its meaning and contextual use. It is often regarded as receptive vocabulary (passive learning), which is known as incidental vocabulary learning (Nation & Waring, 1997), unlike active vocabulary learning which is known as productive vocabulary. Word knowledge and vocabulary are used interchangeably in this study and there is no difference in meaning. Thus, the term receptive knowledge is also known as receptive vocabulary.

Nation (2006) added that the ability of learners to recognise a word and to discriminate a word from words with a similar form or derivatives would mean that they possess the reception vocabulary. Also, it involves the knowledge of judging whether a word form looks and sounds correct as well as predicting the grammatical pattern that suits a word. Productive vocabulary, on the other hand, includes the deeper extension of receptive vocabulary (Nation, 2011). It links to the capabilities to 'productively' or actively apply a word in speech or composition, knowing the correct pronunciation, spelling, and grammatical patterns. In short, receptive vocabulary often involves the ability to make sense of a word while productive vocabulary is linked to the ability to automatically activate a word for dynamic application in various skills. Students enrolled in a specific field need to comprehend the required vocabulary, especially the receptive ones, in their specialised texts to cope with their learning. The measurement of receptive vocabulary is in terms of lexical coverage to determine if the texts are too challenging. When the texts are too challenging, learners, especially second language or foreign language users should spend some time on tackling the unknown but need vocabulary so that revising taking notes from the read texts can be facilitated.

The lexical coverage is "the percentage of running words in the text known by the reader" (Nation, 2006, p.61). Lexical coverage gauges whether a text can be easily or sufficiently comprehended (Laufer & Nation, 1995) and assesses whether a text is adequately learned or comprehended (Hsu, 2014). It is imperative to know how large a receptive vocabulary (recognising a word and discriminating a word from words with a similar form) is needed for daily or typical language use such as reading fiction, academic books, newspapers, engaging in conversations, and even watching a movie. Very often, the quantifiable number of words needed to understand a target material or activity is referred to as vocabulary size. There are various ways to determine the vocabulary size of the vocabulary needed to read and understand texts without external support like using dictionaries. Vocabulary size is measured using the number of word families involved or needed for text comprehension. In terms of morphological knowledge, various definitions may appear where the term 'word family' is concerned. In this study, the research adopts only the definition by Bauer and Nation (1993) as it has been widely used by various researchers (see Nation, 2005; 2011; Hsu, 2009; 2014; Wang, Liang & Ge, 2008; Coxhead & Hirsh, 2007; Chung & Nation, 2003; 2004).

"The most ambitious study in determining vocabulary size carried out tried to work out the total of the number of words there are in English and then set that as the vocabulary learning goal" (Nation, 2006, p.1). Studies were compiling up to 88,500 (Nagy & Anderson, 1984) as well as 114, 000 word families (Goulden, Nation & Read, 1990) and expecting students to master this large vocabulary size is not feasible at all especially in the classroom context. Furthermore, Nation (2006) argued that even first language speakers could not possibly know every word in their first language, and thus, it was insensible to set such a vocabulary learning goal. The ideal scenario has got to be a feasible (minimal) number of vocabulary exposures to achieve the ideal outcome of high lexical coverage in a text. Another way to decide vocabulary learning goals would be to identify what and how much of vocabulary a native speaker knows and make it as a feasible goal. Although there is a list of substantial research in this area, the majority of the methodology used was biased and flawed as the context chosen was either too general or specific (Nation & Hwang, 1995). Nation and Waring (1997) believed that a native speaker scholar would utilise approximately 20,000 word families in their daily application (excluding proper nouns). However, Nation (2006) was concerned about the lack of well-educated research in the area and added that the target of knowing 20,000 word families was over-ambitious. The third way was to determine the vocabulary learning goals of identifying the word families students need to grasp and comprehend reading materials well. For instance, Hsu (2009) explored the vocabulary size needed to adequately comprehend general English textbooks; Coxhead, Stevens, and Tinkle (2010) determined the vocabulary threshold students must have to read science texts, and Hsu (2014) analysed engineering textbooks with the same approach.

Consequently, specialised word lists were developed as suggested lists to be incorporated into vocabulary size analysis (Coxhead, 2000; Coxhead & Hirsh, 2007; Ward, 1999). Previous studies only limited the word lists to those of the initial 2,000 most frequently used words in English (West, 1953) and the University Word List (Xue & Nation, 1984). It was then replaced by the Academic Word List (Coxhead, 2000). Thus, Nation (2006) overcame the limitation and customised the RANGE software (Heatley, Nation & Coxhead, 2002) into the third edition of RANGE software (Nation, 2005). Nation (2005) used the lemma lists from the British National Corpus (BNC), consisting of more than 100-million-word collection - samples of written and spoken language from an extremely wide range of sources. The use of BNC in the third version of RANGE software to compile a substantial 14-word-family list provides a better estimation of the vocabulary size needed for reading and listening (Nation, 2006). Nation (2006) introduced the identification of the size of receptive vocabulary in which students need to grasp any given texts adequately. According to Laufer and Ravenhorst-Kalovski (2010), the most comprehensive approach in determining the needed vocabulary size to read sufficiently is that of Nation's (2006) approach. Trialing 14 frequency lists (1,000 word families per list) in the BNC and checking the coverage of spoken and written texts, Nation (2006) asserted some guidelines. The first 1,000 of the most frequent word families provide coverage ranging from 78 percent to 81 percent in written texts, and an additional eight to nine percent from the second thousand. The third thousand to the fourteenth thousand-word lists would provide less and less incremental lexical coverage.

The vocabulary threshold refers to the vocabulary size needed to comprehend a text at a target percentage. General reading skills can be performed at an ideal level when a reader retains a reasonably massive amount of second language understanding, often termed as the vocabulary threshold. The vocabulary threshold is the minimal vocabulary necessary for "adequate" reading comprehension (Coxhead, Stevens & Tinkle, 2010; Hsu, 2014; Laufer & Ravenhorst-Kalovski, 2010, and Nation, 2006). For instance, if 5,000 word families are determined as the threshold, thus, students are expected to achieve this vocabulary size in order to engage in reading academic materials "reasonably well". Another factor that relates to threshold vocabulary would be "sight vocabulary" (Laufer & Ravenhorst-Kalovski, 2010).

"Sight vocabulary" is defined as words that are very familiar to a reader so that they can be comprehended (recognised and decoded instantly) out of context without any cognitive effort (Laufer & Ravenhorst-Kalovski, 2010; Segalowitz, 2007). For example, if an engineering student encounters a word such as "capacitor" which happens to be in his sight vocabulary capacity in an academic engineering text, he does not need to depend on the other words or context to decode the meaning. Having large sight vocabulary correlates to fluent reading and effortless word recollection to comprehend target content (Segalowitz, 2007). Thus, the related factors of the lexical threshold are the reader's sight vocabulary and lexical coverage which is now defined as the total number of words in the text, excluding real unknown words, and converted into a percentage (Hsu, 2014; Laufer & Ravenhorst-Kalovski, 2010, Nation, 2006). It is claimed that students can comprehend a great number of texts with a comparatively small collection of vocabulary, given the condition that some high-frequency words are frequently repeated. For instance, if a student can acquire the 2,000 most frequently used words in English, the student is expected to comprehend approximately 80 percent of the vocabulary found in common texts (Francis & Kučera, 1982). Hence, the threshold vocabulary would be 2,000 most frequently used words (believed to be sight vocabulary) which cover about 80 percent of lexical coverage in the target text. However, the question that arises is 'what level or percentage of lexical coverage is considered "adequate" or "reasonably well" for a student to read effortlessly and comprehend the cognitive values?'

What is meant by "adequate" or "reasonably well" levels of understanding of a text? It can be very subjective based on the lexical threshold requirement of different contexts. Different educational levels or disciplines may need certain levels of lexical coverage to achieve the desired reading proficiency and even for writing (see Chen, 2020). Furthermore, even in the same discipline, a student might require higher reading standards for a higher academic level. Laufer (1989) was the first to attempt a study to link reading comprehension to lexical coverage by calculating a learners' self-report and underlining unknown words. Results showed that at 95 percent coverage, participants were able to obtain a score of 55 and above. Hu and Nation (2000) also investigated different coverage groups of (80 percent, 90 percent, 95 percent, and 100 percent) to determine the extent of reading comprehension. This study adopts the approach which examines the coverage of different words based on frequency levels provided to texts in specific or representative corpora. Nation's (2006) approach was adopted in this study using the RANGE BNC (Nation, 2005). A similar analysis was carried out only by Hsu (2014) in Taiwan were the texts used were university engineering textbooks.

In the studied corpus, Nation (2006) reported that 8,000 – 9,000 word families (vocabulary threshold) were needed to achieve an ideal lexical coverage of 98 percent. Nation (2006) added that it was possible to reach 98 percent coverage with 5,000 word families and proper nouns as well as 95 per cent coverage with 3,000 word families and proper nouns. The findings were parallel to that of Laufer (1992) who suggested that a 3,000 vocabulary threshold would provide 95 percent lexical coverage. In addition, Laufer (1989), Liu and Nation (1985) have proven that at least 95 percent or higher lexical coverage in a text is required for enhanced success in guessing unknown words. According to Webb and Rogers (2009a), knowing the most frequently used words, including proper nouns, could contribute to comprehending different genres of English television programs. They found that knowing 2,000 to 4,000 word families contributes to 95 per cent comprehension while knowing 8,000 to 9,000 word families contributes to 98 per cent comprehension. In another research by Webb & Rogers on the comprehension of language instructions in movies, it was found that knowledge of 3,000 word families including proper nouns provide lexical coverage of 95 per cent and 6,000 word families provide lexical coverage of 98 per cent (Webb & Rogers, 2009b). As for standardised English proficiency tests like the Test of English for International Communication (TOIEC) and the Test of English as a Foreign Language (TOEFL), Chujo and Oghigian (2009) measured the vocabulary threshold needed for examinees to understand 95 percent of lexical items in the texts. It was reported that an examinee required a range of 4,000 – 4,500 word families to be able to comprehend texts in the tests adequately at 95 per cent lexical coverage (Chujo & Oghigian, 2009). In the case of reading science textbooks, Coxhead, Stevens & Tinkle (2010) showed that to achieve the ideal coverage of 98 percent, a minimum of 11,000 words plus proper nouns threshold vocabulary is required. Also, Hsu (2011) investigated the vocabulary threshold to read core business textbooks in English and found that students needed 3,500 word families (95 per cent lexical coverage) and 5,000 word families (98 per cent lexical coverage). In the same study, Hsu (2011) observed that students needed a higher vocabulary threshold to achieve 95 and 98 per cent lexical coverage in reading business articles (5,000 and 8,000 word families respectively).

However, according to Laufer & Ravenhorst-Kalovski (2010), optimally 8,000 word families would yield a coverage of 98 per cent including proper nouns and 4,000 – 5,000 word families would yield a coverage of 95 per cent including proper nouns, in academic texts. Knowing 95 percent of the words in a text is required for minimal comprehension and successful guessing of words from context (Hu & Nation, 2000; Matsuoka & Hirsh, 2010; Laufer & Ravenhorst-Kalovski, 2010; Read, 2000). Thus, the lexical threshold for successful reading comprehension in this study is set at 95 percent, the minimum percentage needed for text understanding. Hsu (2014; 2009) used the benchmark or threshold of 95 percent as it was believed to be a more feasible goal for her sample.

Likewise, the researcher believes that achieving the lexical percentage of 95 percent should be the ultimate goal for students and teachers considering the vocabulary limitation Malaysian students have (see Kaur, 2013; Mathai, Jamian, & Nair, 2004; Sarimah Shamsudin, Noraini Husin & Amerrudin Abd. Manan, 2013). It is therefore assumed that the absence of the 5 percent lexical coverage in a text can be tolerated, without interfering or affecting the texts comprehension process. Thus, the lexical coverage at the 95 percent level (vocabulary threshold) was set. Only after developing the engineering corpus using the secondary textbooks and running it on the RANGE BNC (Nation, 2005), the researchers can measure the vocabulary size for adequate comprehension of specialised texts. Learning technical content like engineering subjects using English could be overwhelming to learners whose English is a second or foreign language (Ng et al., 2019; Gilmore & Millar, 2018; Thiruchevam et al., 2018; Todd, 2017; Mudraya, 2006). The success of context understanding or text comprehension is often related closely to the vocabulary size of the learners and the extent of lexical coverage of the intended materials in terms of word families. The objective of the current study is to assess the required vocabulary size for the comprehension of engineering texts measured in the percentage of lexical coverage as compared to those of other disciplines at the secondary level. Understanding the vocabulary threshold of certain materials eases course design and lesson planning as well as advancing disciplinary literacy (Green & Lambert, 2018; Bi, 2020).

Methodology

According to Nation (2006), the ideal coverage of vocabulary needed for dealing with any written texts is 8,000 to 9,000 word families. However, in this research, the threshold of vocabulary comprehension input is set at 95 percent based on its feasibility issue (see Hsu, 2014; 2009). Coxhead, Stevens, and Tinkle (2010) in their study of science textbooks in New Zealand reported that to reach the 98 percent lexical coverage, over 11,000 proper nouns word families are needed for Years Nine and Ten and over 15,000 proper nouns are needed for Year eleven texts. The figure opposes the findings found in Nation (2006) as scientific or field-specific texts are different entities. If that is the number of word families needed to be learned by students, it would be too overwhelming and taxing for students and teachers as well. According to Coxhead, Stevens, and Tinkle (2010) analysis, in the range of 4,000 – 9,000 proper nouns, the lexical coverage ranges from 90.96 percent to 97.28 percent in their target science textbooks.

The RANGE BNC software (Nation, 2005) was used to measure the lexical coverage of various textbooks to determine the number of word families needed to comprehend texts. The BNC word family lists in the software consist of 14-word family lists. The 14 lists consist of 1,000 word families selected based on frequency and range data (Nation, 2005). The lists of proper nouns, abbreviations, and marginal words (exclamations and hesitation procedure) were also pre-loaded in the software. These words are believed to be not too difficult for learners and doing without them would inflate the lexical threshold for comprehensible reading. Also, these lists of words were included to analyse the vocabulary demand of the textbooks adopting the research methods used and suggested by Coxhead, Stevens, and Tinkle (2010) Hsu (2009), Kaneko (2015), Nation (2006) and Webb & Rodgers (2009a; 2009b).

The study materials used for the development of the entire textbook collections for secondary 4 and 5 Malaysian vocational schools syllabus are called the *Kurikulum Bersepadu Sekolah Menengah (KBSM),* translated as the Integrated Curriculum Secondary School, engineering textbooks. The texts used were the KBSM Form 4 and 5 engineering textbooks for the field of civil, engineering technology, electrical and electronics as well as mechanical engineering. These textbooks serve as the primary source for Malaysian students who wish to pursue their technical studies early, at the vocational level, especially in engineering. In comparison, the other corpora used to further determine if the engineering texts were too challenging for the secondary school students were compiled. These corpora consist of the KBSM Science series of Form 4 and 5 textbooks - biology chemistry, physics, general science. A non-science related corpus was also compiled, consisting of the KBSM Form 4 and 5 English textbooks. These corpora used for in-depth comparison is referred to as ‘auxiliary corpora’ in this study.

Results and Discussion

To investigate if the 95 percent threshold is achieved, the calculation is demonstrated by summing up the lexical percentage of each BNC word list of the KBSM Civil engineering textbook which is shown in Table 1. The process shown is a sample of how the computation was reached. As for the analysis of the rest of the textbooks, the processes are similar and the data of the number of word families needed to reach the 95 percent threshold for each textbook will be presented in the following tables. Table 1 depicts the overall vocabulary load of the KBSM Civil Engineering textbook (97,227 tokens) with the needed vocabulary size for optimal comprehension. The same method applies to the other three textbooks.

Table 1: Tokens and Lexical Coverage at Each of the BNC Base Word Lists for the KBSM Civil Engineering Textbook

BNC base word lists	Tokens (running words)	Percentage coverage in tokens (%)	Cumulative percentage coverage in tokens	Word families
1st 1,000	741,746	73.14	73.14	985
2nd 1,000	94,520	9.32	82.46	560
3rd 1,000	44,468	4.38	86.84	307
4th 1,000	29,386	2.90	89.74	269
5th 1,000	17,068	1.68	91.42	180
6th 1,000	8,972	0.88	92.3	110
7th 1,000	7,029	0.69	92.99	81
8th 1,000	5,056	0.50	93.49	62
9th 1,000	4,695	0.46	93.95	64
10th 1,000	3,309	0.33	94.28	56
11th 1,000	2,705	0.27	94.55	38
12th 1,000	2,119	0.21	94.76	45
13th 1,000	2,247	0.22	94.98	29
14th 1,000	1,110	0.11	**95.09**	22
Not in the lists	49,777	4.91	100.00	Note
Total	1,014,187	100		7,636

Note: In the RANGE Programme, the words found in the 'Not in the lists' are those that cannot be found in any lists; thus no word families were made for this category.

The first 1,000 word families comprised 66.7 per cent of the tokens and the second 1,000 word families accounted for 12.63 per cent. Thus, the combined coverage of the first two 1,000 word families of the BNC was 79.33 per cent. The figure obtained was much lower than the average 86 to 90 per cent of lexical coverage of various genres as suggested by Nation (2006), showing that students with only 2,000 word families might encounter great challenges when reading the engineering texts. Because of the lack of text familiarity of students would encounter one unknown word while reading the texts. The third highest lexical coverage after the first two word lists was 5.88 per cent, derived from the third 1,000 word families. The coverage percentage of the subsequent 1,000 word families dipped drastically below 4 per cent for the forth 1,000 word families and it rapidly decreased to less than 1 per cent at the sixth 1,000 word families. Only by the eleventh 1,000 word families that the cumulated coverage percentage reached 95.14 per cent (in the bold denomination in

Table 1). It is inferred that ESL engineering students would need a vocabulary of 11,000 word families plus proper nouns, marginal words, and abbreviations to have a vocabulary command of 95 percent of an engineering text. 95 percent is believed to be the minimum vocabulary threshold for acceptable reading comprehension. The alarming fact obtained from the data was that as much as 4.24 percent of the words were not found in any of the word lists, implying that reaching the ideal 98 percent vocabulary threshold as advocated by Nation (2006) was impossible for this textbook. The maximum lexical coverage that could be obtained with 14,000 word families plus proper nouns, marginal words, and abbreviations was only 95.75 per cent. Learning and knowing 14,000 word families in two years is not at all feasible nor practical for ESL engineering students in Malaysia. This task can be too daunting for both teachers and students and this finding coincides with that of Mukundan (2007; 2009) which demonstrates that Malaysian textbooks were developed in an ad hoc manner with irregular vocabulary loading.

The same discussion would follow for the rest of the KBSM engineering textbooks. Based on the BNC word lists scale, the rest of the KBSM engineering textbooks in this study depicted varying vocabulary thresholds. None of the textbooks reached the 98 percent threshold. Hence, reading engineering texts can be of the utmost challenge to students whose first language is not English, especially those who are weak in the language but need to comprehend complex concepts in engineering. The lexical coverage data for all the KBSM engineering textbooks are listed in Table 2.

Table 2: Vocabulary Demands of the KBSM Engineering Textbooks of Four Subject Areas at 95 Percent Lexical Coverage

Specialist Engineering Subject Areas of KBSM Engineering Textbooks	Vocabulary Demands
Civil	11,000
Electrical and Electronics	>14,000
Engineering Technology	11,000
Mechanical	14,000

Note: The vocabulary demand was calculated by counting the number of the BNC 1000-word-family lists necessary plus the proper nouns, transparent compounds, marginal words, and abbreviations until the lexical coverage accumulation reached 95 percent.

As shown in Table 2, reading both the KBSM Civil Engineering and Engineering Technology textbooks requires the least vocabulary demands (11,000 word families at 95 percent lexical coverage). On the other hand, comprehending the KBSM Electrical and Electronics Engineering textbook is the most lexically challenging (more than 14,000 word families at 95 percent lexical coverage). This is followed by the KBSM Mechanical Engineering textbook (14,000 word families at 95 percent lexical coverage). It can be implied that the vocabulary distribution of the KBSM Electrical and Electronics Engineering textbook was the most diverse among the four textbooks, entailing the learning of much larger vocabulary size. Therefore, Electrical and Electronics Engineering students may face more difficulties than those of other majors when they are required to read their specialised textbooks in English. Once again, there is no evidence of standardisation of vocabulary loading in the KBSM Engineering textbooks written by the materials writers across the four engineering subjects.

The vocabulary demands calculation for the KBSM Form 4 and 5 science subjects and English textbooks are similar to those of the KBSM engineering textbooks in terms of the RANGE software used and the method used to reach 95 percent (accumulated) lexical coverage. The vocabulary level analysis of these textbooks (non-engineering) would provide better insights in terms of its direct comparison: whether the engineering textbooks written for students of the same level (upper secondary) were too vocabulary demanding or are just too simplistic for the students. Hsu, in a study conducted in the Taiwan education context (2009), found that the English-medium Business core textbooks required a smaller vocabulary demand as compared to the higher-level general English textbooks. This came as a surprise as field-specific textbooks should demand higher vocabulary command in the students.

In this section, the research seeks to identify the vocabulary demand of the various upper secondary science subjects and English (general) textbooks to determine whether the range of 11,000 to beyond 14,000 word families for the engineering textbooks are too high. This is the first time this type of study and analysis was conducted on Malaysian textbooks to find out the lexical needs of local students. With the data, secondary school teachers, especially for English for Specific Purposes (ESP), can prepare better teaching materials for reading and vocabulary development. The complete data are listed in Table 3.

Table 3: Vocabulary Demands of the KBSM Form Four and Five Science Subjects Textbooks at 95 Percent Lexical Coverage

KBSM Form Four and Five Textbooks	Vocabulary Demands
Biology	14,000
Chemistry	10,000
General Science	9,000
Physics	7,000

The range of word families of the analysed subjects was 7,000 to 14,000, based on the data obtained. The physics textbooks recorded the least number of word families needed for the 95 percent lexical coverage at 7,000 word families. The physics content found is more defined by formulae and diagrams, so that it requires fewer words for comprehension as compared to the other subjects. Therefore, reading physics textbooks may be simple as there are many common or frequent words of English found in the texts. In addition, the smaller number of tokens used in the physics textbooks (the least in this study) contributes to it having the lowest vocabulary demand as the concepts of physics can sometimes be explained through diagrams or formulae. Thus, lesser tokens or vocabulary items were needed to convey information to learners. Like mathematics, physics involves a great bulk of calculations and representations of formulae which do not require a rich command of lexical choices. It can also be implied that generally, the average student does not need an enormous number of lexical items to comprehend physics texts well.

The most lexically demanding subject is biology and this finding coincides with Hsu's (2014) findings that biology-related engineering subject areas were the heaviest in terms of vocabulary loading. It seems that the biology textbooks are the most difficult to read (14,000 word families needed to achieve 95 percent lexical coverage) second only to KBSM Electrical and Electronics textbooks (more than 14,000 word families needed). The vocabulary demand of the biology textbooks is similar to that of the mechanical engineering textbooks. It can be deduced that reading the KBSM Electrical and Electronics textbooks is much more vocabulary demanding than reading the biology and mechanical engineering textbooks. Only by carrying out such comparison, this study can determine the extent of difficulties of the various engineering textbooks. The chemistry textbooks needed 10,000 word families and the general science textbooks 9,000 word families to achieve the vocabulary threshold.

From the results, it can be implied that students who study the subjects of biology, chemistry, physics, and general science on average, need to have approximately 10,000 word families to comprehend relevant texts. On the contrary, vocational school students, who take engineering subjects need on average approximately more than 12,500 word families to cope with reading their texts with optimal understanding. This shows that vocational school students (a mixture of various levels of students) need to have a higher level of vocabulary knowledge than students who take biology, chemistry, physics and general science. This phenomenon proves to be biased and supports the notion that these local textbooks were not properly and systematically streamlined. This error in academic materials' planning makes it even more difficult for the students in vocational schools to succeed in the comprehension of their textbooks.

Table 4: Vocabulary Demands of the KBSM Form Four and Five English Textbooks at 95 Percent Lexical Coverage

KBSM Form Four and Five Textbooks	Vocabulary Demands
English	7,000

Using a similar method discussed above, the vocabulary demand of the KBSM Form Four and Five English Textbooks was determined. This analysis was needed because it could provide a better comparative study to find out the lexical difficulty of general English texts. It provides a benchmark of the vocabulary needed in reading general texts of a similar education level (upper secondary). As listed in Table 4, it did not come as a surprise that the vocabulary demand for English was the lowest among all with only 7,000 word families needed for 95 percent comprehension of the texts which is similar to that of the physics vocabulary demand. Although reading general English texts requires the same number of vocabulary demand with that of physics texts, the vocabulary demand to the tokens ratio of the general English texts is much higher. This may suggest that reading general English texts can be more challenging than reading physics texts. From comparison with previous data, the findings suggest that reading specialised texts (except the physics texts) requires more than just general English vocabulary. Hence, students in the field of engineering need to pay extra attention to the specialised or field-specific academic vocabulary to cope with the lexical challenges, as the average vocabulary demand in the engineering texts is on average more than 12,500 word families.

Conclusion

The text coverage threshold value set is 95 percent text comprehension for this study. From the analysis carried out, for the engineering texts, especially in the electric and electronics upper secondary engineering textbooks pose the highest vocabulary demand of 15,000. This does not come as a surprise as this discipline is believed by many to be one of the toughest disciplines in engineering alone. On average most of the auxiliary textbooks used as comparative analysis show lower vocabulary demand except the subject of biology. The subjects of physics and English record the lowest vocabulary demand. Knowing the vocabulary demand of each textbook is crucial to students so that the textbooks will not be too overwhelming for them. Material writers can use this type of analysis to derive suitable words to be introduced in the textbooks and control the selection more systematically. Also, proper standardisation of the introduction of lexical items can be facilitated with the support of such data. In other words, if certain textbooks are found to be overly lexically challenging to students, the syllabus or content of the textbooks should be revamped. In this manner, the textbooks can be fully utilised by students should they feel that they have a high level of understanding or at least the grasp of the content when they know most or even up to 90 percent of the words. Ideally, achieving the 95 percent target should be aimed; however, it means that students need to put in more effort and time in acquiring the needed words. Having some sort of guideline or a series of word lists to follow would ease this process of acquiring the relevant vocabulary as well as some help with imagery of 'vocabulary viewing' (Durbahn, Rodgers, & Peters, 2020).

Funding

This work was supported by the **BOLD 2025 Grant Funding Universiti Tenaga Nasional (Project Code – 10436494/B/2019050)**

Acknowledgment

The researchers are sincerely thankful for the support and fund provided by iRMC, Universiti Tenaga Nasional to carry out this study under the BOLD 2025 Grant Funding Universiti Tenaga Nasional *(Ref: "Developing Energy Literacy through Language in the 21st Century Learning Project Code – 10436494/B/2019050)*

REFERENCES

Bauer, L. & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279

Bi, J. (2020). How large a vocabulary do Chinese computer science undergraduates need to read English-medium specialist textbooks? *English for Specific Purposes, 58*, 77-89.

Chen, Y. (2020). Accessing the lexical richness of figurative expressions in Taiwanese EFL learners' writing. *Assessing Writing, 43*, 100414.

Chujo, K. & Oghigian, K. (2009). How many words do you need to understand TOEIC, TOEFL & EIKEN? An examination of text coverage and high frequency vocabulary. *The Journal of Asian TEFL, 6*(2), 121 – 148.

Chung, T. M. & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*, 103-116.

Chung, T.M. & Nation, I.S.P. (2004). Identifying technical vocabulary. *System, 32*(2), 251-263.

Coxhead, A. & Hirsh, D. (2007). A pilot science word list for EAP. *Revue Française de Linguistique Appliqueé XII* (2), 65-78.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Coxhead, A., Stevens, L. & Tinkle, J. (2010). Why might secondary science textbooks be difficult to read? *New Zealand Studies in Applied Linguistics, 16* (2), 37-52.

Durbahn, M., Rodgers, M. & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System, 88*, 102166

Francis, W.N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, USA: Houghton Mifflin.

Gilmore, A. & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes 51*, 1–17

Goulden, R., Nation, P., & Read, J. (1990) How large can a receptive vocabulary be? *Applied Linguistics, 11*(4), 341–363.

Green, C. & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes, 35*, 105-115

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). RANGE and FREQUENCY programs. Available online: http://www.vuw.ac.nz/lals/staff/Paul_Nation

Hsu, W. (2009). Measuring the vocabulary of college general English textbooks and English-medium textbooks of business core courses. *Electronic Journal of Foreign Language Teaching, 6* (2), 126-149.

Hsu, W. (2011). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes, 30*(4), 247–257.

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes, 33* (1), 54-65.

Hu, M. & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.

Kaneko, M. (2015). Vocabulary size required for the TOEFL iBT listening section. *The Language Teacher, 39*(1), 9-14.

Kaur, N. (2013). A case study of tertiary learners' capability in lexical knowledge. *GEMA Online Journal of Language Studies, 13* (1), 113-126.

Laufer, B. & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30.

Laufer, B. and Nation, P. (1995). Lexical richness in L2 written production: Can it be measured? *Applied Linguistics, 16* (3), 307-322.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp.126–132). Basingstoke & London: Macmillan.

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.

Liu, N. & Nation, I.S.P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal, 16* (1), 33-42.

Mathai, E.J., Jamian, L.S. & Nair, S. (2004). Assessing Malaysian university students' English vocabulary knowledge. In W. Khanittanan & P. Sidwell (Eds.), *SEALSXIV: Papers from the 14th meeting of the Southeast Asian Linguistics Society Volume 1* (pp. 219-237). Canberra: Pacific Linguistics.

Matsuoka, W. & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT coursebook provide good opportunities? *Reading in a Foreign Language, 22*(1), 56–70

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purpose, 25*, 235-256. doi:10.1016/j.esp.2005.05.002

Mukundan, J. (2007). Irregularities in vocabulary load & distribution in same level textbooks written by different writers. *Indonesian JELT 3*(1), 99- 118.

Mukundan, J. (2009). *ESL textbook evaluation: A composite framework*. Köln, Germany: Lambert Academic Publishing.

Nagy, W.E., & Anderson, R.C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*(3), 304–330.

Nation, I. S. P. (2005). *The BNC word family lists 14,000.* Retrieved from http://www.victoria.ac.nz/lal/about/staff/paul-nation.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59–82.

Nation, I.S.P. & Hwang, K. (1995). Where would general service vocabulary stop & special purposes vocabulary begin? *System, 23* (1), 35-41.

Nation, I.S.P. & Waring, R. (1997). Vocabulary size, text coverage & word lists. In N. Schmitt & M.McCarthy (Eds.), *Vocabulary: Description, Acquisition & Pedagogy* (pp. 6-19). Cambridge, UK: Cambridge University Press.

Nation, I.S.P. (2011). New roles for L2 vocabulary. In V. Cook & L.Wei (Eds.), *Contemporary Applied Linguistics: Language Teaching and Learning Volume 1.* (pp. 99-116). London, UK: Continuum International Publishing Group.

Ng, Y.J., Chong, S.T., Loh, K. S. Ahmad Zufrie Abd Rahman, Thiruchelvam, S. (2019). Developing a Vocational Engineering Word List: Bridging the Gaps of Secondary and Tertiary Education. *International Journal of Advanced Science and Technology, 28*(10), pp. 104-112

Read, J. (2000). Assessing vocabulary. Cambridge, UK: Cambridge University Press.

Sarimah Shamsudin, Noraini Husin & Amerrudin Abd. Manan (2013). Exploring fundamental engineering word list for engineering students: A literature review. *Procedia - Social and Behavioral Sciences*, 70, 1275-1281. doi:10.1016/j.sbspro.2013.01.187.

Segalowitz, N. (2007). Access fluidity, attention control and the acquisition of fluency in a second language. *TESOL Quarterly, 41*, 181 – 186.

Thiruchelvam, S., Ng, Y.J, Chong, S.T., Azrul Ghazali & Norhayati Bte Mat Husin (2018). The Language of Civil Engineering: Corpus-based Studies on Vocational School Textbooks in Malaysia. *International Journal of Engineering & Technology, 7* (4.35), 844-847

Todd, R.W. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes, 45*, 31-39.

Wang, J., Liang, S.L. & Ge, G.C. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*, 442-458. doi:10.1016/j.esp.2008.05.003

Ward, J (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language, 12*(2), 309-323.

Webb, S. & Rodgers, M.(2009a). Vocabulary demand of television. *Language Learning, 59* (2), 335 - 366.

Webb, S. & Rodgers, M.(2009b). The lexical coverage of movies. *Applied Linguistics, 30* (3), 407 - 427. doi:10.1093/applin/amp010

West, M. (1953). *A general service list of English words*. London, UK: Longman.

Xue, Guoyi, & Nation, I. S. P. (1984). A university word list. *Language Learning & Communication, 3*(2), 215–229.