MEASURING LUBOK ANTU'S YEAR SIX PRIMARY SCHOOL PUPILS' VOCABULARY USE

Mohamad Hafizuddin bin Ma'mor
Sekolah Kebangsaan Ridan
Lubok Antu, Sarawak
pikachu_pika89@yahoo.com

Hamidah binti Yamat
Faculty of Education
Universiti Kebangsaan Malaysia
hamidah_yamat@ukm.edu.my

ABSTRACT

Vocabulary plays an integral part in measuring one's language ability to comprehend and produce language. Poor performance in English examinations may have a direct relationship with vocabulary deficiency especially for primary school pupils in a rural district such as Lubok Antu in Sarawak. With the lack of vocabulary research in Malaysia especially in a rural district context, there is a dire need for a vocabulary research to be carried out. This research aims to measure the vocabulary use by Year Six pupils in Lubok Antu. A total of 116 Year Six pupils from seven different rural level schools participated in this study. This study measured pupils' vocabulary use through their two written essays from their Ujian Penilaian Sekolah Rendah (Primary School Assessment) trial writing paper. All the essays were checked and words that were irrelevant were removed before each word was typed and analysed through RANGE program. The produced output illustrated pupils' vocabulary use in terms of word frequencies and the words' distribution among word frequency levels. The results showed that pupils had no difficulty in producing words to meet the word limits for both essays but they had a high word repetition rate. They were inclined to use a number of word family in their written works thus exemplifying a possible vocabulary deficiency that might cause limited use of a variety of vocabulary in their writings. This research forms a sound background data and a foundation for further extended research to figure out the actual vocabulary use of pupils in a rural district's primary schools thus prompting extra care has to be taken before a change or action is implemented to help intervene and improve the current issue.

Field of Research: *vocabulary, English language teaching, rural district, Sarawak*

----

1. Introduction/Background

Studies on English language learners' corpus have provided the background to more studies documenting English learners' vocabulary knowledge, competency or use. A range of Malaysian English language learners' corpus are also documented from pupils' English corpus to corpus which is specifically based on certain genres (Abdul Rahim, 2014). That study cites several English learners' corpus studies that are local such as Abdul Samad et al.'s (2004) English of Malaysian School Students (EMAS), Knowles and Zuraidah's (2005) Malaysian Corpus of Learner English (MACLE), Botley et. al's (2005) Corpus Archive of Learner English Sabah-Sarawak (CALES) and other corpus of learners which are specific on certain genres like Business and Management English Language Learner Corpus and the Engineering Lecture Corpus (ELC). Throughout all these learners corpus' studies, there is a plenty of rooms to further develop those studies in a more specific or a wide-range context. This is crucial especially for a rural and less-researched district in Malaysia such as Lubok Antu, Sarawak. A study as such, will contribute to pedagogical perspectives that informs real vocabulary use input for English teachers, English for teaching and learning material developers and curriculum designers in Malaysia. The vocabulary lists from this research will serve as a vocabulary

baseline and assist practitioners to make more informed decisions especially for pupils in a rural district like Lubok Antu, Sarawak.

This current study discussed in this article centres on a chief issue which is pupils' inefficiency in English reading more so writing. Based on a study by Chang, Lin and Mohamed (2014), it is found out that eleven year old pupils in Malaysian rural primary schools have poor reading comprehension skills. The same situation is apparent in Lubok Antu, Sarawak where other subjects that are tested in *UPSR* outperform English subjects thus reflecting a deficiency in terms of pupils in Lubok Antu's English proficiency in answering English *UPSR* examination papers. A high average point shows underperforming average points which denote that many pupils do not get the minimum passing grade for that paper. The data are tabulated in Table 1 below.

Table 1: *Gred Purata Mata Pelajaran UPSR* (*UPSR* Average point for examination subject) for Lubok Antu's district

Year	*Gred Purata Mata Pelajaran UPSR* (Average point for examination subject) (out of 5.00)				
	Malay Language (Comprehension)	Malay Language (Writing)	Science	Mathematics	English
2011	2.45	2.20	2.74	2.71	3.24
2012	2.31	2.31	2.66	2.67	3.00
2013	2.12	2.31	2.73	2.37	2.90
2014	2.09	2.23	2.54	2.52	2.88
2015	2.22	2.36	2.44	2.57	2.92

This trend supports the research earlier by Chang, Lin and Mohamed (2014) and this scenario paints a clear picture that specifically, rural primary school pupils may not have appropriate reading skills to comprehend English. Ultimately, they are not able to produce English output orally or in written form. It is widely acknowledged that reading is another main input or receiving method from learners apart from listening to a language. If pupils have difficulties in comprehending listening tasks, poor vocabulary might be the possible cause of the comprehension difficulty. The same assumption could be applied in reading as pupils' poor vocabulary may be the cause of reading incomprehensibility. Thus, these issues will become critical problems as these pupils are at disadvantaged. It is apparent that there is a need to explore pupils' actual vocabulary use.

The current research aims to measure the vocabulary use by Year Six pupils in Lubok Antu, Sarawak, Malaysia. This research focuses on a main objective which is to analyse Year Six pupils' writings to measure their vocabulary use through their writings based on two writing sections of their English writing *UPSR* trial examination papers. The unit of analysis is based on two guided written English essays of Lubok Antu Year Six primary school pupils from different rural levels. All the essays are gathered and processed before using a computer program to analyse and compartmentalise all words according to word counting mechanisms (word tokens, word types and word families) and word frequency levels of wordlists. The literature and methodology are reviewed and explained beforehand before the research's findings are reported and discussed. The implications and suggestions from this research are provided to further rectify this research issue.

2. Literature Review
2.1 Vocabulary assessment

It is the primary nature of this research to assess pupils' vocabulary thus providing ways to tap into their vocabulary knowledge. There are two parts of assessing pupils' vocabulary. One, is the common vocabulary size testing assesses learners' receptive/passive vocabulary on written vocabulary. In terms of vocabulary size, Nation and Gu (2010) describe two tests to assess learners' vocabulary size which are Vocabulary Levels Test and Vocabulary Size test. Both Vocabulary Size Test and Vocabulary Levels Tests are evaluating learners' receptive knowledge. However, these tests

require pupils' active participations thus making them impractical for rural schools' pupils to participate.

The other part of vocabulary assessment is productive vocabulary tests. In terms of actual language production, there are two possible ways to gather learners' language production which are oral production and learners' written productions. Getting learners' written productions is more practical as Read (1993) views oral interviews are time consuming. Thus, the current study focuses on learners' written works to get an understanding of their vocabulary knowledge. Productive vocabulary test offers a platform for vocabulary evaluators to assess learners' actual language production hence complementing receptive vocabulary tests. For productive vocabulary assessment, learners' written works are assessed in terms of its vocabulary composition and its relations to the word frequency levels. To do that, learners' words and their learning materials in their written compositions have to be counted systematically. Nation and Gu (2010) proposes lexical richness measure as a measure to assess learners' written works although it might not portray learners' actual vocabulary size as it provides a platform to assess learners' ability to use their vocabulary knowledge communicatively. They describe three plausible ways to count words in learners' writings and learning materials through word tokens, word types and word families. All three counting methods differ from each other based on several counting criteria. For word tokens, words are counted at every instance although the words have their repetitions and this differed from word types where words are tallied at each instance but repetitions of the same word are not taken into account for the whole data set. As it is suggested from its label, words that will be counted through word families consider their headword, inflections and derivatives from the words thus only the headword is counted. Table 3 below displays the examples of the differences of counting words between word tokens, word types and word families. For word frequency levels, it had different layers from high frequency words to low frequency words and academic vocabulary and technical vocabulary (Nation & Gu, 2010).

In Malaysia, there is still a lack of research on assessing primary school pupils' vocabulary knowledge aside for Abdul Samad et al's English of Malaysian School Students (as cited in Abdul Samad, 2004). Although it is reported that studies that are related to corpus of English locally has its roots in 1990's (Abdul Rahim, 2014), it is either difficult to access or hardly any research that has attempted to document the vocabulary use of the pupils similar to this study's focus especially in Sarawak. This research specifically opts to extend Abdul Samad et. al's (2004) English of Malaysian School Students (EMAS) and Ma'mor and Md Yunus' (2015) study as its research backgrounds. Although both studies have different objectives, there are similar shared research elements that are suitable for knowledge expansion. It is time for more data from various contexts and situations, especially in a district in Sarawak, to be gathered and analysed so that a clearer picture of the actual primary school pupils' performance in vocabulary is documented. This research's findings can help improve materials like textbook as several research point out deficiencies in Malaysian ELT textbooks (see Singh & Jerry, 2008; Zarifi & Mukundan, 2012; Khostajeh & Kafipour, 2012; Mukundan & Khostajeh, 2011; Mukundan & Roslim, 2009; Mukundan & Hussin, 2006). It is also worrying if the trend is continuing without any necessary actions are taken to improve the situation. There is a need for continuous evaluation of textbooks so that its lacks are rectified and textbooks will be improved (Zohrabi, 2011). Undoubtedly, the provided textbook has a direct link to the pupils' vocabulary knowledge and vocabulary use thus forming a cycle. Actual pupils' vocabulary use could help improve textbook because if the provided textbook has been shown to have its lapses, it would be possible that textbook deficiencies could affect pupils' poor reading comprehension skills and eventually leading to their poor vocabulary use.

3. Methodology
3.1 Research design, sampling and participants
This research utilised a descriptive study design with a quantitative focus. This will eventually forms a small-scale corpus of pupils' vocabulary use through measuring their written essays. 16 out of 27 Lubok Antu's Malay medium national schools were purposely selected because of the schools' willingness to participate in this study including several that are rurally located. The schools that returned their pupils' examination papers were eight schools. This made the return rate 50%. However, a school's sample was not analysed due to the incorrect examination paper. Thus, the actual sample, consisted of seven schools with a total of 116 English writing examination papers, was received. This research's 116 pupils were around 25% of Lubok Antu's 473 Year Six pupils.

3.2 Instrument
This research utilised a written test as its main instrument. The written test was taken from the current English writing examination paper which the second English paper of *Ujian Penilaian Sekolah Rendah* (Primary School Assessment). It consisted of three sections. The first section was the information transfer. The second section was a short 60 to 80 words formatted written essay and the third section consisted of an 80 to 100 words picture stimulus essay where pupils choose one picture stimulus to write on from two picture stimuli. This research adopted the second and the third section and not the first section due to the first section's limited written vocabulary production in that section as the chances of pupils to use their own words were slim as they were expected to just copy the given information and might not exploit or add other words in their written production. By having the last two sections of the writing paper, pupils' vocabulary production might be more flexible and the possibility of getting pupils to utilise their vocabulary knowledge in production is higher. It is practical and appropriate to use the available test instrument because all schools in the research's population have used the same writing test paper for their *UPSR* trial examination.

The main analysing tool for this research was Range. Nation and Gu (2010) describe Range software as a platform that utilises General Service List and Academic word List to produce Lexical Frequency Profile (LFP). LFP is a concept and a tool on its own that is introduced by Laufer and Nation (1995) and they explain LFP has the ability to place and provides percentages of learners' produced written texts based on frequency levels that vary. Range program will produce its output in terms of the material vocabulary use and the material's lexical profile. In addition, its offline nature makes it suitable for people who are interested to use it who may not have the Internet privilege especially in this research's rural context. Another reason is that Range is used previously in EMAS corpus and a research on concordance based on EMAS corpus by Abdul Samad (2004). In terms of reliability and availability, Laufer and Nation (1995) assert that vocabulary use in written works that are assessed in terms of LFP is considered valid.

3.3 Data collection procedure and data analysis
The first stage of this research's content analysis was the editing process where the participants' written works were edited before each word from the collection of all the participants' written works were typed in a .txt format file. It is acknowledged that the editing process focuses on serious word errors and not grammatical errors as this research focused on vocabulary use and not grammatical properties of the participants' written works. Serious word errors include poor use of words in the wrong context and irrelevancy. Laufer and Nation (1995) argue that incorrect word use reflect the absence of the word in one's word knowledge. All the punctuations are removed during the editing process as the presence of punctuations will affect Range output. Range treats some punctuation as command codes. However, spelling errors that did not distort the meaning of the sentences were accepted and repaired and this may indicate that the participants may not have fully mastered the correct spelling of the words but they use the words correctly to fit the sentences' meaning and context.

After the editing process, the second stage was the typing stage. Each participant's written works were typed in a single .txt file continuously with the rest of the participants as the total words of the participants' words in their written works were processed together through Range. The third stage was the .txt file that was run through Range to produce a LFP of each focused material. For this research, the wordlists that were utilised by Range was The General Service List (GSL) and Academic Word List (AWL) and not BBC corpus list. Nation and Gu (2010) point out that GSL consists of the 1000 and 2000 high frequency words that are regular from West and vocabulary that is used for formal academic purposes which is AWL from Coxhead. In Malaysia, the same wordlists' combination (GSL and AWL) is adopted by Abdul Samad et al.'s (as cited in Abdul Samad 2004) research on Year Five's written works. Thus, it is a well-researched combination and this ensures the consistency of this research's data analysis. With the chosen wordlists, Range produced its output in terms of word frequencies of the word, token types and word families with word frequency levels thus forming the basis of the research's findings.

4. Findings & Discussion

Table 2: Pupils' written vocabulary use

Tokens			Types			Families		
19062 (100%)			957 (100%)			425		
Base List One	Base List Two	Base List Three	Base List One	Base List Two	Base List Three	Base List One	Base List Two	Base List Three
14019 (73.54%)	1668 (8.75%)	73 (0.38%)	484 (50.57%)	109 (11.39%)	7 (0.73%)	334	87	4
Not in the list 3302 (17.32%)			Not in the list 357 (37.30%)			Not in the list X		

A total of 19062 words were produced from the combination of two essays from each pupil of the 116 pupils in this study. Thus, each pupil wrote an average of 164 words in two essays (82 words for each essay). This is a promising sign as both essay questions have a word limit of 180 words for the combination of both essays' word limits and the minimum word limit is 140 words for both essays. The total token count shows that pupils in this study have the ability to write as near as the total word limits for both essays. This may be viewed positively because pupils in this study are able to fulfil the minimum word limits for both essays despite their constraints. From the word token counts, it can be seen that pupils can produce the number of words that are required for them to answer the questions' minimum word limit requirements but that does not fully display their writing ability in terms of utilising words in their writings. This could be illustrated by looking at the word types from pupils' writings. As compared to word tokens' concept that counts every single word in pupils' writings, word types puts a counting concept where the repetitions of a single word are not counted. Only 957 word types are detected from the whole 19062 word tokens. Type-token ratio that compares the percentage of type to the overall tokens is used here. The type-token ratio for pupils' vocabulary use is around 5% and this shows a rough estimation of a high level of word repetitions in pupils' written works as 5% of the total word tokens are word types. There is dramatic decrease of words in word types from its word tokens. 95.5% of the total word tokens which around 18105 words of pupils' writings are repetitions from 957 word types. This shows no doubt pupils have manage to write more than the minimum word limits for both essays but it is clear that they tend to recycle their words in their written works. This could lead to a notion that they tend to use words that are familiar to them. They keep repeating the words in their written works just to fulfil the provided tasks and this may show that pupils' inability to use new or other varied words.

To verify the above notion, the word families of the total word tokens that are produced by pupils are scrutinised. It is noticed that only 425 word families that were detected from the total word tokens and word types. Only the root form of the word or the base word is counted in word families distinction thus overlooking the root form's other multiple other forms such as its word derivations and inflections. 44% words from the word types are consisted of words in word families and these word families form a mere 2% form the word tokens in pupils' writings. As word families focuses on the word's basic form, it lends more supports that the actual collection of base words from pupils' written works is based on a small group of base words. This means that the pupils' vocabulary list or knowledge is limited.

Next, it is interesting to note further breakdowns of each word token, word type and word families' distinction in terms of its word frequency level. This step is important as it is intriguing to assess each word counting area and its relations to the high frequency words, academic vocabulary and others. First, the biggest word area that carries the most weight is the word tokens show that 73.54% (14019 words) of the total word tokens are from the first one thousand high frequency words. Progressing to the second one thousand high frequency words, 8.75% or 1668 words are detected in this word frequency band and 0.38% (73 words) is from the base list three which is the academic vocabulary list. On the surface level, the overall word tokens form pupils' written works are primarily based on the first one thousand high frequency words from West's General Service List. It is noted that 17.32% (3302 words) is not in these three lists and this could be attributed to the presence of proper nouns and words that are low in frequency as compared to the focused three frequency lists. Then, the percentages decreases dramatically when the word frequency level moves from frequent to less frequent thus denotes the first one thousand frequency words' prominence in pupils' written works.

The same trend could be observed when the word frequency level breakdown is applied to pupils written works' word types and word families. For word types, 484 word types (50.57%) detected are in the first one thousand high frequency words are detected as compared to 109 word types (11.39%) which consisted of the second one thousand high frequency words. Only seven word types (0.735) from the academic word list are present out of a total of 957 word types in pupils' written works. 357 word types (37.30%) make up the total word types and this figure could be reasoned with the presence of proper nouns that pupils' use and possible presences of low frequency words. To substantiate the trend, pupil written works' word families is analysed. It is seen that 334 word families are used from the first one thousand high frequency words, 87 word families from the second one thousand high frequency words and four word families that are listed in the academic word list. There is no word families distinction for words out from the first three word lists as it would be taxing to argue a basic form for each proper nouns such as people's and places' names.

The findings of the current study are then compared to two similar local studies which are namely Abdul Samad et al's EMAS corpus (as cited in Abdul Samad, 2004) and Ma'mor and Md Yunus' study (2015). The comparison here is primarily for added knowledge and information as these similar studies have different objectives than the current study. Firstly, Abdul Samad et al's EMAS corpus (as cited in Abdul Samad, 2004) has corpus data on Year Five primary school pupils apart from secondary school students' data although this research has different sample size and the number of collected pupils' written works. EMAS corpus is chosen to be benchmarked and discussed due to the absence of published research articles that are similar to this study even if only some of its findings could be integrated such as average word length and the distribution of words in word frequency levels or wordlists for this study's closest research reference. Abdul Samad (2004) reports that EMAS corpus has a collection of 82218 words with the inclusion of the proper nouns from a collection of three essays from each pupil in that study. It is calculated that a pupil in EMAS corpus study has written an average of 93 words in an essay from his or her three essays. This is comparable

to this study's finding where pupils in this study have produced an average of 82 words in an essay from two essays that have been gathered. Although it clearly shows that pupils in this study has lower average word count from EMAS corpus study that might hint their vocabulary deficiency, Abdul Samad (2004) describes EMAS study as a study that utilises written works from above average pupils. It is highly likely that above average pupils in EMAS corpus study could write more words than this study's rural district pupils even though they are one year younger. However, the gap between produced written words of this study and EMAS corpus study is eleven words which is not distant and this demonstrates that pupils in the current study perform considerably well despite their conditions.

In comparison to Ma'mor and Md Yunus' study (2015), their study does not provide the total word count for their study as they only state the total word count (3518 words) without the proper nouns. The total word count should be higher when the proper nouns are included. These proper nouns do impact in the total word count thus automatically affecting the average word count for a pupil's single essay but with the proper nouns do not have a direct effect in terms of word frequency levels or wordlists as they are categorised in the not-in-the-list's list. In Ma'mor and Md Yunus' study (2015), it is found that these urban-located Year Six pupils have written an average of 113 words per essay. Similar to EMAS corpus study, this study does not inform any word limit for each pupil's written production thus increasing the probability that pupils could produce more words without worrying about stated word limit like in this study. The current study does not focus on a specific ability group of pupils like in the mentioned two studies thus reflecting a mixture of ability in terms of English writing skills and this study has employed written tasks that have word limits. Interestingly, pupils in this study could produce up to an average of 90 words per essay due to the word limits for both essays as there are two scenarios that have happened. First, pupils write beyond the word limits which are 180 words for the combination of both essays. The second scenario is some pupils write considerably less thus affecting the average word count of the whole study's finding. Nonetheless, it does not hinder the fact that there is still a gap between pupils' vocabulary use in this study with pupils' vocabulary use in these two similar studies especially in terms of the average written word count in a single essay.

Next, word frequency levels or wordlists data from EMAS corpus study and Ma'mor and Md Yunus' study (2015) provide another comparative dimension to the discussion. The comparison can provide insights on the frequency of produced words in terms of word usage in the first one thousand high frequency words, second one thousand high frequency words and academic word list. Although it is arguable that EMAS corpus has a bigger sample size and higher number of collected pupils' written works and Ma'mor and Md Yunus' study (2015) has only 35 pupils' written works from 35 pupils , it is motivating to observe how pupils in this study match up to those two studies. Table 3 and Table 4 below illustrate the comparison between this study's findings and those two studies based on word frequency levels.

Table 3: Comparison between this study's findings and those two studies based on word frequency levels in terms of word types.

Types								
First one thousand high frequency words			Second one thousand high frequency words			Academic word list		
This study	EMAS	Ma'mor and Md Yunus (2015)	This study	EMAS	Ma'mor and Md Yunus (2015)	This study	EMAS	Ma'mor and Md Yunus (2015)
484	770	301	109	250	57	7	29	10

Table 4: Comparison between this study's findings and those two studies based on word frequency levels in terms of word types.

Families								
First one thousand high frequency words			Second one thousand high frequency words			Academic word list		
This study	EMAS	Ma'mor and Md Yunus (2015)	This study	EMAS	Ma'mor and Md Yunus (2015)	This study	EMAS	Ma'mor and Md Yunus (2015)
334	470	220	87	180	47	4	25	10

It could be observed that these three studies show a consistent declining pattern based on their word frequency levels' data for both word types and word families. Pupils in all three studies have produced words primarily in the first one thousand high frequency words and steadily decreasing in use for the next consequent wordlists. The current study follows very much of the same decreasing pattern from the other two similar studies thus supporting those study's findings. Additionally, this is consistent with the declining pattern from a study by Engku Ibrahim, Sarudin and Muhamad (2016) where they find out that the majority of pre-university students' receptive vocabulary of first and second one thousand high frequency words is at the highest before it declines for the subsequent low frequency wordlists. This supports the notion that pupils' vocabulary use in this study is stable even if it is compared to a study which is focusing on older participants and utilising receptive vocabulary which has constant bigger size than productive vocabulary according to Laufer et al (2004). Therefore, the current study's findings are stable when the figures are matched up to studies which focus on both receptive and productive vocabulary.

The declining pattern based on word frequency wordlists from frequent to less frequent does not the paint whole picture due to the comparison of this study to two studies who have both extremes. EMAS corpus study has bigger samples and Ma'mor and Md Yunus' study (2015) has much smaller samples. Unsurprisingly, EMAS corpus study surpasses the current study and Ma'mor and Md Yunus' study (2015) in terms of the number of words in each word frequency levels or wordlists for both word types and word families due to its bigger sample size and higher level of English proficiency. However, it is seen that pupils in Ma'mor and Md Yunus' study (2015) do not fall far in terms of produced words in the first and second high frequency word levels when it is compared to the current study. Ma'mor and Md Yunus' study (2015) has samples that are 30% out of the current study's samples, shares the same age and schooling level which is Year Six and still, their study's samples have vocabulary use that do not vary much from the current study. Although Ma'mor and Md Yunus' study (2015) employ pupils from the first class out of two classes that might denote their samples consist of average to above average English proficiency pupils, the current study has a mixed-ability pupils and it is safe to say that a number of the pupils in the current study are average and above average. This is because the current study's vocabulary use has more words in the first and second high frequency words than Ma'mor and Md Yunus' study (2015) but they could produce far more different words due to their number of samples. Remarkably, that is not the case for academic word list. Ma'mor and Md Yunus' study (2015) has more words in academic word list than the current study. This demonstrates a gap especially in academic word list where pupils in the current study could produce more words in the first and second high frequency word lists but they are still behind in producing words in academic word list.

Based on the comparisons for word frequency levels between the current study and the two similar studies, it is similar with the average word count before where the current study's vocabulary use can be improved based on the number of pupils in the study. This is because EMAS corpus study outperforms the current study with a doubled higher word count in nearly each word level of word types and word families although EMAS corpus study's samples are Year Five pupils who are younger

than the current study's samples. The age different might balance out the fact that EMAS corpus study has above average pupils and the sample size is more than a double from the current study. The fact that the sample size is doubled, it would be logical if the current study's vocabulary use to match up to EMAS corpus' vocabulary use. When the current study is compared to a study which has smaller sample size, the vocabulary issue is apparent as much smaller sample study is able to replicate the same situation with the comparison of EMAS corpus study with the current study where the current study's vocabulary use is just a double of number of word types and word families in the majority of word frequency levels than Ma'mor and Md Yunus' study (2015). Pupils in this study tend to recycle words in their writings and do not produce different word types thus affecting their word families' count. Furthermore, the vocabulary issue is far apparent when it is equated to the sample's actual 2016 *UPSR* examination results for their English writing paper. The average subject grade or *Gred Purata Matapelajaran (GPMP)* for the seven participating schools is 3.45 out of 5.00 that demonstrate that the average pupils' grade for English writing paper is in the C and D grade which are just satisfactory. Thus, it is possible that the pupils' vocabulary issue has had an impact in their actual examination result and this study could validate and provide justification to the released examination result. All and all, there is a need for pupils to start using more varied words to construct sentences and convey meanings in their writings thus providing opportunities for them to improve lexically.

5. Implications and Conclusion

This study's pedagogical implications are based on its pupils' corpus. Although this study's corpus is still relatively small, this corpus has the potential to be exploited as a valuable input for material development that is suitable for primary schools in rural districts. This study lists frequently-used words by pupils in this study's context and it serves as a guideline for word choice to be used in English teaching and learning materials that are current and future. The current material that can be improved through the use of the frequent-words' list is the provided English textbook. By having pupils' frequent-used words, teachers who share similar environment with this research are able to use to the list to do a quick-check of the face value of the textbook or any reading texts. This will increase these teachers' awareness towards words that hold the highest probability to be comprehended and eventually will be used by pupils. Moreover, teachers could replace words that are deemed 'difficult' such as words beyond the first and second one thousand high frequency words to words from this study's frequently-used words. These steps would increase the probability for pupils to comprehend the teacher-provided reading texts for classroom teaching and reading modules thus transforming them to be more suitable based on pupils' vocabulary data. As well, pupils' vocabulary use in this study offers an insight for future material developers especially for textbook authors and English supplementary material producers. This study's findings could form a vocabulary benchmark for these material developers when their materials are targeted for pupils who are specifically in Sarawak's rural districts. This would assist in the production of English materials that are better suited for pupils who are similar to this research's context.

In conclusion, the data from this study provide insightful information for different aspects of English teaching and learning's materials from adapting and adjustments for current and future English materials. This study paves ways to more vocabulary friendly English materials through more similar research to measure pupils' actual vocabulary use in different situations and contexts in the future.

References

Abdul Rahim, H. (2014). Corpora in language research in Malaysia. *Kajian Malaysia, 32*(1), 1-16. Retrieved from www.academia.edu/download/33759358/Corpora_in_Language_Research_in_Malaysia.pdf

Abdul. Samad, A. (2004). Beyond concordance lines: Using concordances to investigating language development. *Internet Journal of e-Language Learning & Teaching, 1*(1), 43-51. Retrieved from www.abdelali.net/ref/archad-beyond.2004.pdf

Chang, K. L., Lin, S. E., & Mohamed, A. R. (2014). Benchmarking Year Five students' reading abilities. *English Language Teaching, 7*(5), 50-58. Retrieved from http://www.ccsenet.org/journal/index.php/elt/article/download/36053/20296

Engku Ibrahim, E. H., Sarudin, I. & Muhamad, A. J. (2016). The relationship between vocabulary size and reading comprehension of ESL learners. *English Language Teaching, 9*(2), 116-123. Retrieved from http://www.ccsenet.org/journal/index.php/elt/article/viewFile/56388/30189

Khojasteh, L., & Kafipour, R. (2012). Are Modal Auxiliaries in Malaysian English Language Textbooks in Line with Their Use in Real Language? *English Language Teaching, 5*(2), 68-77. Retrieved from www.ccsenet.org/journal/index.php/elt/article/viewFile/14564/9929

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307-322. Retrieved from https://www.researchgate.net/profile/Batia_Laufer/publication/249237693_Vocabulary_Size_and_Use_Lexical_Richness_in_L2_Written_Production/links/544689af0cf22b3c14de3a94.pdf

Laufer, B., Elder, C., Hill, K. & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing, 21*(2), 202-226. Retrieved from http://www.academia.edu/download/36306849/2004_size_and_strength_-_do_we_need_both_to_measure_vocabulary_knowledge.pdf

Ma'mor, M. H. & Md Yunus, M. (2015). Utilising a corpus of Malaysian national primary school pupils' English writings in ELT material selection. *Jurnal Penyelidikan Pendidikan 2015, 16*, 157-169.

Mukundan, J. & Hussin, A. A. (2006). Automatic and semi-automatic process of Wordsmith 3.0 as a textbook evaluation instrument: A case study. *TEFLIN Journal, 17*(2), 200-219. Retrieved from http://www.journal.teflin.org/index.php/journal/article/download/68/65

Mukundan, J. & Roslim, N. (2009). Textbook representation of prepositions. *English Language Teaching, 2*(4), 13-24. Retrieved from files.eric.ed.gov/fulltext/EJ1083708.pdf

Mukundan, J., & Khojasteh, L. (2011). Modal auxiliary verbs in prescribed Malaysian English textbooks. *English Language Teaching, 4*(1): 79-89. Retrieved from www.ccsenet.org/journal/index.php/elt/article/download/9668/6916

Nation, P. (1993). Using dictionaries to estimate vocabulary size: Essential, but rarely followed, procedures. *Language Testing, 10*(1), 27-40. Retrieved from http://ltj.sagepub.com/content/10/1/27.short

Nation, P. & Gu, P. Y. (2010). *Focus on vocabulary.* Victoria, Australia: Macmillan.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355-371. Retrieved from http://ltj.sagepub.com/content/10/3/355.short

Zarifi, A. & Mukundan, J. (2012). Phrasal verbs in Malaysian ESL textbooks. *English Language Teaching, 5*(5), 9-18. Retrieved from www.ccsenet.org/journal/index.php/elt/article/download/16650/11104