

Measuring Khmer Vocabulary Size for Each Grade in Basic Education

BUNLOTKHOY

Faculty of Social Sciences and Humanities
Royal University of Phnom Penh
Phnom Penh, Cambodia

Abstract:- People communicate each other in uttering any meaningful way with vocabulary. The knowledge of vocabulary and the ability to access that knowledge efficiently are recognized as important factors in reading and listening comprehension and speaking and writing fluency. Thus, vocabulary knowledge helps students with language comprehension. The purpose of this study has been to determine the measuring Khmer vocabulary size for each grade in basic education. Quantitative method was adopted to gather primary data from two newspapers issued for a week and nine books of folktales. Descriptive statistics were then used to investigate the frequencies of words that were used in the one-week issue of the two daily newspapers and in a set of nice Khmer folktale story books. The results indicated that the mean of grand total of sources as the mean of the number of titles was ($M = 45$ and $S.D = 15$) while the mean of non-repeated words was almost ($M = 3000$ and $S.D = 500$) and the mean of repeated words was closely ($M = 20000$ and $S.D = 5000$) using a whole article per day or an entire book. Furthermore, the result showed that each grade in basic education was approximately for 1,000 new words based on the frequency usage dividing into six different categorical word frequencies. The results of this study are helpful for curriculum developers and publishers to create a wordlist of Khmer vocabulary for basic education.

Keywords:- Basic Education, Vocabulary, Vocabulary Size, Word Frequency.

I. INTRODUCTION

According to the Ministry of Education, Youth and Sport (MoEYS) (2015a), Cambodia's education system is divided into four levels. First is kindergarten. This level allows students to attain school from the age of three to five. The kindergarten is categorized into three ranges such as low, medium and high. Secondly, it is the foundation of education for students from the age of six to 14, equivalent to the first grade to the ninth grade. This level is divided into primary education (grade 1 to 6) and lower secondary education (grade 7 to grade 9). The third level is upper secondary school. The students can decide whether to study the upper secondary education (grade 10 to 12) or the Technical and Vocational Education and Training (TVET) (vocational and technical Level 1 to 3). This level is for students aged 15 to 17. The final level is higher education

for students aged 18 or over. In this level, students can study Associate, Bachelor, Master, and PhD degrees. Moreover, TVET's students have trained in Associate, Bachelor, Master, and doctorate degrees of the specialized technical expertise. Thus, the education system in Cambodia includes kindergarten, foundation of education, upper secondary education and higher education.

Correspondingly, the Ministry of Education, Youth & Sport (MoEYS) has paid close attention in the tightening of the grade 12 national exam since academic years 2013-14 and since then successfully completed grade 12 students with grade A, B, or C can automatically become teacher trainees in primary school (MoEYS, 2015b). What is more, the curriculum reform has been undertaken at all levels from early childhood education to Technical and Vocational Education and Training (TVET) and high education to ensure that important and applied knowledge, demanded skills and professional and ethical attitude will be imparted to current and future Cambodian students. The grade 12 national exams and the curriculum reform as well as the upgrading of teachers' qualification and capability have been given priorities to improve instruction at all school levels (MoEYS, 2015b).

MoEYS (2015b) and World Bank's (2014) recommendations on how to improve quality and capable teaching force in Cambodia can be categorized into five main themes: (1) recruiting potential candidate teachers to be trained in Pre-Service Trainings (PRESET); (2) generating conducive environment for teaching and learning; (3) ensuring effective and efficient evaluation and monitoring; (4) upgrading teachers' qualifications, pedagogical content knowledge (PCK) and professional development In-Service Trainings (INSET); (5) providing competitive compensation in forms of salary, other incentives, good working condition and better social status as well as welfare.

Some researchers and MoEYS focused on improving the quality the grade 12 national exams, rebuilding the curriculum from basic education and TVET to high education, promoting qualified teachers, and inserting dictation in grade 9 national exams. Inappropriately, Khmer vocabulary sizes mainly plays an important role in leading the effective and efficiency learners both receptive and productive skills. But it is insignificant interested and not yet properly standardized in sound and spelling, let alone a

list of words to be learned by grade in order of level of difficulty and frequency. Such an underdevelopment presents challenges for the learners and users. Mistakes in writing is everywhere from a few lines in social media to books. Furthermore, an assessment report in Khmer language for grade 3 and 6 students showed that less than 50% of them who took the test scored the passing rate (MoEYS, 2017). Contributing factors can be many. One of which is the list of systematic analysis of the vocabulary to be taught by lower secondary and how they should be taught. This study will attempt to count and analyze the frequency of the words from two sources as media and the 9 books of folktales, using Dictionary of Khmer Language and Cambodian Dictionary Part I & II as the main standard for counting words and word formations. Therefore, critical questions that cannot be ignored and which this study endeavors to address are as follows:

1. How many Khmer vocabulary word frequencies were used in the two daily newspapers per week and one set of Khmer folk tale story books?
2. How many and what words in each grade are for Khmer vocabulary of basic education based on frequency of its usage, in comparison to news sources?

II. LITERATURE REVIEW

2.1 Definition of Vocabulary

Vocabulary has been defined differently by different researchers. Burns and Broman (1975) referred to “the stock of words used by a person, class, or profession” (p. 259). Hubbard (1983) argued that “vocabulary can be defined as a powerful carrier of meaning, 67”. Hatch & Brown (1995) stated that the term vocabulary refers to a collection or set of terms which may be used by individual language speakers for a particular language or a series of words. Hornby (1995) described vocabulary as the total number of words in a language; a list of words with their meanings is vocabulary. Cameron (2001) defined vocabulary as one of the linguistic information fields which plays a major role for learners who acquire a language. Daimond and Gutlohn (2006) stated that vocabulary is the knowledge of words and word meaning. Neuman and Dwyer (2009) referred to vocabulary as “The total number of words that are needed to communicate ideas and express the speakers' meaning” (p.385).

2.2 The Importance of Vocabulary

Krashen and Terrell (2000) suggest that in order to convey their meaning, language learners must be able to produce lexical objects if they wish to express some concepts or ask for information. According to Schmitt (2000), vocabulary in language teaching indicated that vocabulary is central to communicative competence and to the acquisition of a second language. Vocabulary and lexical subjects are the core elements of learning and communication. Without vocabulary mediation, no amount of grammatical or other forms of linguistic knowledge can be used in conversation or discourse since vocabulary is shown to be much more concentrated than the knowledge of single words. In communication situations, Davies and Pearse (2000) point out that vocabulary is often more important than grammar. Language learners are frustrated as they find that they

cannot communicate effectively because they don't know many of the words they need. This is also reiterated by McCarthy (1990) who points out the value of vocabulary that no matter how well the student learns grammar, no matter how well the sounds of L2 are mastered, with no words to express a wide range of meanings, communication in an L2 simply cannot take place in any meaningful manner.

McKeown (2002) claimed that knowledge of vocabulary is at the heart of a language understanding and use. Vocabulary competence is one of the most important factors influencing language learning, and it is understood that vocabulary information helps improve language learning (Martin-Chang & Gould, 2008). Hence, if learners read more, they can develop their knowledge of vocabulary and vocabulary instruction improves not only the pronunciation of the learners but also their writing skills. According to Harmer (2007), learners were constantly tested in vocabulary learning to find out how they can develop their level of vocabulary knowledge. If learners do not have sufficient vocabulary knowledge, they will have difficulties in recognizing the content, thereby failing the vocabulary learning test and losing interest in reading.

2.3 Measuring Vocabulary Size

A vocabulary size test measures just how many words a learner understands. It typically measures the knowledge of the word form and the ability of a learner to link that form to a meaning. A measure of receptive vocabulary size looks at what kind of knowledge is needed for listening and reading. A receptive test measures whether the learners can give or choose meaning when seeing the word type. The finding of Kirkpatrick (1891) presented that the vocabulary of native speakers of English at between 10,000 and 100,000 words dependent on education. Thorndike and Lorge's (1944) results indicated that learners knew around 17,000 basic words and they indicated an average learning rate of just under 1,000-word families. Seashore and Eckerson (1947) suggested college students learned more than 58,000 basic words.

Additionally, Hartmann (1946) produced estimates of 200,000 words known by undergraduates. Lorge and Chall (1963) estimated there were 16,900 basic words known to first graders. Diller (1978) estimated that over 50,000 words were known to seven-year-olds, and more than 100,000 words to junior high school students and children learn 20,000 words per year at the secondary school. Nagy and Herman (1984) estimated of 40,000 words known by college seniors. Miller and Gildea (1987) recommended that school students learn 5,000 words of vocabulary per year. White, Graves, and Slater (1990) suggest 60,000 words to educated monolingual speakers. Goulden, Nation, and Read (1990) estimate the vocabulary size of a well-educated adult native English speaker in research for L1 learners as subjects to be around 17,000 words.

According to D'Anna et al. (1991), undergraduate students know 16,785 words by recognition and 14,076 words which they could define. In addition, the receptive vocabulary exposure of English native speakers grew by up

to twenty years at the rate of 1,000-word families per year and also indicated that awareness of 5,000-word families is essential to enjoy reading(Hirsh & Nation, 1992).Laufer and Nation (1995)stated that a text coverage of 95% could be reached with a 5,000-word English vocabulary. Biemillerand Slonim(2001) found an average vocabulary size of 5,200 root words (word families) for six-year old which increased to 8,400 root words by nine years old. Such statistics roughly corresponded to adding two or three words to the responsive vocabulary of the children every day, resulting in an increase of about 1000-word families per year.

Biemillerand Boote(2006) estimated the number of known root meanings at the end of second graders to be around 6000, rising by 1000 per year and addressed the hypothesis and stated that American second grader adds 1,000 new meanings of words per year. Nation (2006) asserted that 8,000 to 9,000-word families are needed for understanding a written text and a vocabulary of 6,000 to 7,000-word families for comprehension of spoken text, if 98% coverage of a text is desired. Nonetheless, there was one statistic that is repeated regularly and has acquired the status of being a normal, and this is the number that kids learn about 3,000 new words a year (Graves, 2006).Milton (2009) proposed a common language for British undergraduates as limited as 9,000- word families. Milton and Treffers-Daller(2013)found UK undergraduates enter university with a 10,000-word definition vocabulary and gain about 500 words each year while studying.

2.4 Word Frequency

A suitable measure of word helpfulness is frequency, which is showed how often a word occurs in normal use of a language orlarge proportion of both written and spoken texts that occur very frequently. It makes sense that language learners are taught words which belong to the most frequent lexemes of the language when those frequent words are likely to be the most useful. It is consistent with Laufer and Nation (1999)who argued thatteachers of English should

focus on the 2000 most frequent words, on high-frequency vocabulary, and instead of teaching individual words which are less frequent, they should conduct strategies for deal with unfamiliar vocabulary. They have conducted several studies which, unsurprisingly, found that high-frequency words are mastered better than low-frequency words.

Stæhr(2008)stated that the 2000 most frequent English words constituted an important watershed and a sensible learning goal in lower secondary school students in Denmark, as those learners who knew those frequent words also performed relatively well in listening, reading and writing.The researcher employed the same tool as that ofSchmitt, Schmitt and Clapham(2001) to measure receptive vocabulary sizes. They clarified that the academic word level from the test was not relevant for low-proficiency learners.Similarly, the finding of Nation (2011)showed thathigh- and low-frequency vocabularies are frequency based with the standard of high-frequency vocabularyincluding 2,000 most frequent word families. Plus, 80% of any fixed English text covered by high-frequency vocabulary.

III. METHODOLOGY

3.1 Data Source for the study

The data for this study is based on articles collected from two sources, i.e., two popular online newspapers in Cambodia (as indicated by numbers of daily site visitors and shares on social media such as Facebook) and 9 books of Khmer folktales (some of Khmer folktales including Khmer curriculum). The articles were issued all topics that are archiving for the study one week after they were published on the two newspapers’ official websites in July 2017 and 9 books of folktales for this study. In total, 831 articles and 248 titles of Khmer folktales were collected. Therefore, all articles and Khmer folktales were selected. Table 3.1 shows the number of articles and titles of Khmer folktales analyzed in the study by online newspaper and 9 books of folktales.

Table 3. 1*Total number of articles collected listed by newspaper and titles of Khmer folktales*

Name of sources	Number of stories or titles
<i>Raksmei Kampuchea Daily (one-week issue)</i>	443
<i>Kohsantepheap Daily (one-week issue)</i>	388
<i>9 books folk talk stories</i>	248
Grand total	1079

3.2 Research Design and Instrument

The researcher adopted quantitative method to collect data to answer the research questions. In this quantitative method, the researcher employed two tools in order to describe the present practice ofthe measuring Khmer vocabulary size for each grade in basic education.Therefore, the instrument would be used to obtain the primary data (831 newspaper articles and 9 books of folktales) for this study, which were then analyzed in the form of content and statistical analysis using PAN Khmer Line Breaking Program by using the standard of Dictionary of Khmer

LanguagePart I & II and word count tools would be utilized to identify the word frequency. Furthermore, the Excel program was used to remove duplicates and sort and filter words and Cambodian Dictionary Part I & II would be applied to be classified by word formation.

3.3 Data Collection Procedure

The researchers checked the official websites of Rasmei Kampuchea and Kohsantepheap to retrieve all articles and 9 books of Khmer folktales issued in the timeframe identified above. The researchers searched for

articles by using the date of issues through the entire websites. All articles were copied and pasted on to the Microsoft word program for archiving as a text database for analysis.

3.4 Data Analysis and Statistical Procedure

After collecting all articles and Khmer folktales, the researchers used the PAN Line Breaking Program to separate words by zero space. The researcher then used the “find and replace” in Microsoft word program by using the Paragraph Mark to count the words. No text in the pictures, headlines, captions, name of authors, proper nouns, and date of publication were counted and analyzed. The data for this study was analyzed using the computer software program as Excel. Word count in this study conducted in word counter tools. Descriptive statistics were utilized to analyze the data for this study. Descriptive statistics provide information on minimum, maximum, frequency distribution, mean, and standard deviation to count words.

IV. RESEARCH FINDINGS

From the findings for this study are divided into two sections. The first section analyzed the word frequencies from two diary newspapers and one set of Khmer folk talk stories using PAN Khmer Line Breaking Program and relied on the standard of Dictionary of Khmer Language and Word count tools for the first research questions. The second section analyzed in the form of content and statistical analysis using the same program and standard to identify the word frequency. Furthermore, the Excel program was used to remove duplicates and sort and filter words and Cambodian Dictionary Part I & II were used to be classified by word formation for the second research question. Before starting with analyzing research questions, the demographic data related to 9 books of folk talk stories and the one-week issues of Raksmei Kampuchea and Kohsantepheap Dailies were presented in Table 4.1 which presents a summary of the demographic sources of the titles and stories which provided the valid vocabulary.

Table 4.1: Outline of all data sources

Sources	Titles	Non-repeated words	Repeated words
9 books of folktales	248	8,528	261,276
Raksmei Kampuchea daily (one-week issue)	443	6,046	136,127
Kohsantepheap daily (one-week issue)	388	6,518	100,131
Grand Total	1,079	9,712	495,235

Table 4.1 presented the demographic variables of vocabulary. There was a total of 1,079 titles with 9,712 non-repeated words and 495,235 repeated words which were analyzed in this study. The most title newspaper of Raksmei Kampuchea daily for a week were among 443 titles to compare with 388 titles of Kohsantepheap daily for a week and 248 titles of 9 book folk talk stories while 9 book folk talk stories were the most 8,528 non-repeated words with 261,276 repeated words then Kohsantepheap daily for a week was 6,518 non-repeated words more than Raksmei Kampuchea daily as 6,046 non-repeated words. However, Raksmei Kampuchea daily was 136,127 repeated words more than Kohsantepheap daily consisting 100,131 repeated

words. Thus, the result showed that the more repeated words increased, the more non-repeated word increased.

4.1 Results for the word frequency from the three sources

The first research question examined in this study was: “How many the frequency words were used in the two daily newspapers per week and one set of Khmer folk tale story books?” This question was to explore the vocabulary word frequencies from the three data sources. The aim is to find out whether the vocabulary word frequencies from each book for folk talk story and each day of Raksmei Kampuchea and Kohsantepheap newspapers.

Table 4.2: Outline of 9 books folk talk stories

Titles of folk talk stories	Number of stories	Non-repeated words	Repeated words
Part 1	34	2,979	38,027
Part 2	23	3,397	47,653
Part 3	53	1,912	24,619
Part 4	12	3,624	24,767
Part 5	20	3,546	30,613
Part 6	20	3,575	27,325
Part 7	25	2,886	19,195
Part 8	29	3,408	24,046
Part 9	32	3,522	25,031
Grand total	248	8,528	261,276

As showed in the table 4.2, the 9 book of folk talk stories consisted of 248 stories with 8,528 non-repeated words and 261,276 repeated words. Each part of the folk talk stories ranges from the minimum number of stories, 12, to the maximum number of stories, 53 with the lowest vocabulary of 1,912 non-repeated words to the highest vocabulary of 3,624 non-repeated words and the lowermost

number of 19,195 repeated words to the uppermost number of 47,653 repeated words. The minimum number of stories part 4 compared with maximum number of stories part 3 different among five times while non-repeated words of part 4 were much more two times of non-repeated word of part 3. The repeated words of part 4 were the comparable size on the repeated words of part 3.

Table 4. 3: Outline of Mean of 9 books folk talk stories

9 book folk talk stories	Minimum	Maximum	Mean	SD
Number of titles	12	53	28	12
Non-repeated words	1912	3624	3205	551
Repeated words	19195	47653	29031	8720

As seen in the table 4.3, the outline of mean of grand total of 9 book of folk talk stories were covered with number of titles, non-repeated words, and repeated words. The mean of number of titles was (M = 28 and S.D = 12) while mean of non-repeated words was (M=3205 and S.D = 551) and mean

of repeated words was (M=29031 and S.D = 8720). The findings showed that an average book of the folk talk story was uttered approximately 3000 non-repeated words and 30000 repeated words.

Table 4. 4: Outline of Raksmei Kampuchea Daily (one-week issue)

Day issued	Number of titles	Non-repeated words	Repeated words
Day 1 for Monday	58	2,214	15,151
Day 2 for Tuesday	78	2,520	21,661
Day 3 for Wednesday	76	2,443	20,504
Day 4 for Thursday	83	2,920	20,577
Day 5 for Friday	52	2,320	13,725
Day 6 for Saturday	35	1,863	9,422
Day 7 for Sunday	61	2,487	15,289
Grand total	443	6,046	136,127

As displayed in Table 4.4, the Raksmei Kampuchea Daily (one-week issue) contained for 443 number of titles with 6,046 non-repeated words and 136,127 repeated words. The minimum number of titles was issued on Saturday that consisted of 35 titles with 1,863 non-repeated words and 9,422 repeated words while the maximum number of titles was issued on Thursday that was 83 titles with 2,920 non-repeated words and 20,577 repeated words. The lowest non-

repeated words were then on Saturday that included 1,863 words while the highest repeated words were on Thursday that was among 2,920 words. Moreover, the lowermost repeated words were on Saturday including 9,422 words whereas the uppermost repeated words were on Tuesday covering 21,504 words. Thus, Reksmei Kampuchea Daily was written the most common article issues on six days a week, but the lowest common on Saturday.

Table 4. 5: Outline of Kohsantepheap Daily for a week

Day issued	Number of titles	Non-repeated words	Repeated words
Day 1 for Monday	64	2,582	19,342
Day 2 for Tuesday	63	2,746	19,692
Day 3 for Wednesday	65	2,481	17,453
Day 4 for Thursday	52	3,131	23,632
Day 5 for Friday	56	2,865	17,737
Day 6 for Saturday	24	1,427	6,779
Day 7 for Sunday	64	2,103	11,456
Grand total	388	6,518	100,131

As shown in Table 4.5, the one-week issue of the Kohsantepheap Daily contained for 388 number of titles with 6,518 non-repeated words and 100,131 repeated words. The outmost number of titles were on Monday, Tuesday, Wednesday, and Sunday of Kohsantepheap Daily, which are about 65 titles while the number of titles was on Friday, which is 56, and the lowermost number of titles was on Saturday, which is 24 titles. Over 2500 non-repeated words was written on Monday, Tuesday, Thursday, and Friday

whereas under 2500 non-repeated words was uttered on Saturday and Sunday. Plus, the uppermost repeated words were on Thursday, which is more than 20000 words, while the bottommost repeated words were on Saturday, which is under 10000 words, and between 10000 and 20000 repeated words were on Monday, Tuesday, Wednesday, Friday, and Sunday. Thus, the result showed that weekday was the most common written, while the weekend day was the lowest common utterance in producing articles for issues.

Table 4. 6: Outline of Mean of Kohsantepheap and Raksmei Kampuchea Dailies for a day

No	Two Newspapers	Minimum	Maximum	Mean	SD
Kohsantepheap					
1	Number of titles	35	83	63	17
2	Non-repeated words	1863	2920	2395	322
3	Repeated words	9422	21661	16618	4477
Raksmei Kampuchea					
1	Number of titles	24	65	55	15
2	Non-repeated words	1427	3131	2476	563
3	Repeated words	6779	23632	16584	5649
Grand total of the two sources					
1	Number of titles	30	71	59	14
2	Non-repeated words	1645	3026	2436	420
3	Repeated words	8101	22105	16601	4768

As shown Table 4.6, the outline of mean of Kohsantepheap was the mean of number of titles (M = 63 and S.D =17) while mean of non-repeated words was (M=2395 and S.D = 322) and mean of repeated words was (M=16618 and S.D = 4477). Moreover, the outline of mean of Raksmei Kampuchea was the mean of number of titles (M = 55 and S.D =15) while mean of non-repeated words was (M=2476 and S.D = 563) and mean of repeated words

was (M=16584 and S.D = 5649). Furthermore, the outline of mean of grand total of two sources was the mean of number of titles (M = 59 and S.D =14) while mean of non-repeated words was (M=2436 and S.D = 420) and mean of repeated words was (M=16601 and S.D = 4768). Therefore, authors of two newspapers as Kohsantepheap and Raksmei Kampuchea, which each one is almost 2000 non-repeated words per day, regularly wrote.

Table 4. 7: Outline of Mean of Grand total of Sources per day

Grand total of sources	Minimum	Maximum	Mean	S.D
Number of titles	26	65	44	13
Non-repeated words	2279	3522	2845	467
Repeated words	14509	29669	21920	4765

As Table 4.7 shows, the outline of mean of grand total of sources was the mean of number of titles (M = 45 and S.D =15) while mean of non-repeated words was (M=3000 and S.D = 500) and mean of repeated words was (M=20000 and S.D = 5000) using whole article per day. Hence, both book and newspapers were commonly written approximately 3000 non-repeated words and 20000 repeated words, which are issues per day or a book publication.

4.2 Results for the frequency and type of words in each grade

The second research question examined in this study was: "What extent is how many and what words in each grade for Khmer vocabulary of basic education based on frequency of its usage, in comparison to news sources?" This question was to investigate the vocabulary of basic education and how many and what words in each grade. The aim is to find out the frequency and type of words included in the basic education from each grade including classified word formation such as (1) consonant + vowel; (2) consonant + vowel + consonant; (3) consonant + sub-consonant + vowel + (consonant); (4) consonant + consonant + sub-consonant + (sub-consonant) + vowel + (consonant); (5) Pali and Sanskrit.

Table 4. 8: Classification of word formation

Khmer word formation	Khmer words	Number of words
consonant + vowel	កិច្ចៈខើរកៅអី	908
consonant + vowel + consonant	កោងមងវិកាយ	2,513
consonant + sub-consonant + vowel + (consonant)	ស្តាំប្រះឆ្ងៃងក្នុង	2,443
consonant + consonant + sub-consonant + (sub-consonant) + vowel	បង្កើរង្វេងចង្រ្កោះសង្រ្កាម	998
Pali and Sanskrit	កម្មករចិត្តវិន័យព្រឹត្តិការណ៍	2,580
Grand total		9,712

As showed in Table 4.8, Khmer word formation was divided into five dimensions that c ranged from 908 words to 2,580 words and 9,712 grand total words. Plus, Khmer word formation was 908 words as consonant + vowel; 2,513 words as consonant + vowel + consonant; 2,443 words as consonant + sub-consonant + vowel + (consonant); 998 words as consonant + consonant + sub-consonant + (sub-

consonant) + vowel; and 2,580 words as Pali and Sanskrit. After receiving 9,712 words from three sources, the grand total words were divided into nine as fundamental educations in Cambodia that starts from first grade to ninth grade. The result showed that each grade was about for 1,000 new words based on frequency usage in chronological order.

Table 4. 9: 9,000 words compares with news sources per article and chapter

Compare with news sources	Repeated	Non-repeated	% (compare with 9,000 words)
ThmeyThmey News	1,471	476	0.93%
Camnews	2,491	576	0.94%
Resmei Kampuchea News	1,576	587	0.91%
Fresh News	1,210	412	0.96%
Sabay News	1,052	387	0.90%
Khmer Literature grade 12	2,549	750	0.93%
History in higher education	1,458	451	0.89%
Study guidebook for high school students	1,611	412	0.96%
Critical thinking book for teachers	1,686	434	0.95%
Grand total	15,105	1,790	0.93%

As Table 4.9 shows, in order to compare 9,000 words as mentioned inserting to basic educational curriculum to news sources such as ThmeyThmey, Camnews, ResmeiKampuchear, Fresh News, and Sabay newspapers per article and Khmer literature book for grade 12, history books for higher education, study guidebooks for high school students, and critical thinking books for teachers per chapter. The results showed that grand total overlapped words were 93%. The findings showedthat 9,000 words for basic educational words for putting in curriculum was scientific evidence to suggest that a guidance for curriculum developers and publishers.

grand total of 9 book folk talk stories was the mean of number of stories, which is 28, while mean of non-repeated words was close to 3000words and mean of repeated words was similar to 20000 words. It may be authors of two online newspapers and a series of 9 books of Khmer folktales to use the level of contextual, ritual, and limited words in narrative the text. Moreover, Khmer language may be likely to update words slowly and no level of words to organize from the most to rarest frequency usage. Thus, authors wrote an article and a Khmer folktale book around the mean of 3000 non-repeated words and 20000 repeated words, which are issues per day or a book publication.

V. DISCUSSION AND CONCLUSION

This research study tried to examine “measuring Khmer vocabulary size for each grade in basic education” It tried to answer two questions. As explained in the findings, there was a total of 1,000 titles with 10,000 non-repeated words and 500,000 repeated words analyzed in this study. The outline of mean of Kohsantepheap and Raksmei Kampuchea dailies for one-week issue was 60 average of titles with the mean of 2,500 non-repeated words and 20000 repeated words for a day issued and the outline of mean of

Regarding theword frequency, the result showed that each grade was approximately 1,000 new words based on frequency usage in a chronological order. The vocabulary was divided into five dimensions of word formations ranging from 908 words to 2,580 words and 9,712 grand totalwords. Additionally, Khmer word formation was 908 words as consonant + vowel; 2,513 words as consonant + vowel + consonant; 2,443 words as consonant + sub-consonant + vowel + (consonant); 998 words as consonant + consonant + sub-consonant + (sub-consonant) + vowel; and 2,580 words as Pali and Sanskrit.It may be likely to use

10,000 non-repeated words to verify with the mean of Kohsantepheap, Raksmei Kampuchea, 9 book of Khmer folktales which were utilized daily and it may progressively increase 3 times of the mean of non-repeated words. It is assumed that readers have an opportunity to read the books or newspapers easily if they know approximately 10,000 non-repeated words and that students use 1,000 non-repeated words for first grade to 9,000 words in ninth grade, so students might read and write words effectively and efficiently from the books and newspapers. The results from this study are consistent with the findings of Hirsh and Nation (1992) found that English native speakers grow by up to twenty years at the rate of 1,000-word families per year and also indicate that awareness of 5,000-word families is essential to enjoy reading. Biemiller and Slonim (2001) and Biemiller and Boote (2006) stated resulting in an increase of about 1000-word families per year. Nation (2006) asserts that 8,000 to 9,000-word families are needed for understanding a written text and a vocabulary of 6,000 to 7,000-word families for comprehension of spoken text, if 98% coverage of a text is desired. Milton (2009) proposes a common language for British undergraduates as limited as 9,000-word families. Milton and Treffers-Daller (2013) indicated UK undergraduates enter university with a 10,000-word definition vocabulary and gain about 500 words each year while studying. The results from this study are inconsistent with the findings of Hartmann (1946) produced estimates of 200,000 words known by undergraduates. Then, the finding of Diller (1978) and Kirkpatrick (1891) presented that the vocabulary of native speakers of English at between 10,000 and 100,000 words dependent on education. Plus, Seashore and Eckerson (1947), Miller and Gildea (1987), and White, Graves, and Slater (1990) suggest 60,000 words to educated monolingual speakers. Nagy and Herman (1984) estimated of 40,000 words known by college seniors. Furthermore, Thorndike and Lorge's (1944), Lorge and Chall (1963), Goulden, Nation, and Read (1990), and D'Anna et al. (1991), estimate the vocabulary size of a well-educated adult native English speaker in research for L1 learners as subjects to be around 17,000 words.

To conclusion, this study is attempted to the measuring Khmer vocabulary size for each grade in basic education. Based on the results of the study the mean of grand total of sources was the mean of number of titles ($M = 45$ and $S.D = 15$) while mean of non-repeated words was ($M = 3000$ and $S.D = 500$) and mean of repeated words was nearly ($M = 20000$ and $S.D = 5000$). Furthermore, the result showed that each grade was approximately for 1,000 new words based on frequency usage starting from first grade to ninth grade divided into six different categories. The findings showed that 9,000 words for basic educational words for putting in curriculum was scientific evidence to suggest a guidance for curriculum developers and publishers. Thus, the results of this study should be considered in the positive effectiveness on Khmer vocabulary for basic education and developing on wordlist of a guidance for curriculum developers and publishers or this result of the study should be discussed for further studies.

REFERENCES

- [1]. Biemiller, A., & Boote, C. (2006). An effective method for building vocabulary in primary grades. *Journal of Educational Psychology*, 98(1), 44–62. <https://doi.org/10.1037/0022-0663.98.1.44>
- [2]. Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, 93(3), 498–520. <https://doi.org/10.1037/0022-0663.93.3.498>
- [3]. Burns, C. P., & Broman, L. B. (1975). *The language arts in childhood education*. Chicago: Rand McNally College Publishing Company.
- [4]. Cameron, L. (2001). *Teaching languages to young learners*. Cambridge: Cambridge University Press.
- [5]. D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a Meaningful Definition of Vocabulary Size. *Journal of Reading Behavior*, 23(1), 109–122.
- [6]. Davies, P., & Pearse, E. (2000). *Success in English teaching*. Oxford: Oxford University Press.
- [7]. Diamond, L., & Gutlohn, L. (2006). *Vocabulary handbook*. Berkeley, CA: Brookes.
- [8]. Diller, K. C. (1978). *The language teaching controversy*. Rowley, Mass: Newbury House.
- [9]. Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(1), 341–363. <https://doi.org/10.1093/applin/11.4.341>
- [10]. Graves, F. M. (2006). *The Vocabulary Book*. New York, NY: Teachers College Press.
- [11]. Harmer, J. (2007). *How to teach English*. Essex, England: Pearson Education Limited.
- [12]. Hartmann, G. W. (1946). Further evidence on the unexpected large size of recognition vocabularies among college Students. *Journal of Educational Psychology*, 37(1), 436–439.
- [13]. Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language education*. New York, NY: Cambridge University Press.
- [14]. Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(1), 689–696.
- [15]. Hornby, A. S. (1995). *Oxford advanced learners' dictionary of current English*. Oxford: Oxford University Press.
- [16]. Hubbard, P. (1983). *A training Course for TEFL*. Oxford: Oxford University Press.
- [17]. Kirkpatrick, E. (1891). The number of words in an ordinary vocabulary. *Science*, 18(446), 107–108.
- [18]. Krashen, S. D., & Terrell, T. D. (2000). *The natural approach: Language acquisition in the classroom*. New York, NY: Longman.
- [19]. Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(1), 307–322. <https://doi.org/10.1093/applin/16.3.307>

- [20]. Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33–51. <https://doi.org/10.1177/026553229901600103>
- [21]. Lorge, I., & Chall, J. (1963). Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *Journal of Experimental Education*, 32(2), 147–157. <https://doi.org/10.1080/00220973.1963.11010819>
- [22]. Martin-Chang, S. Y., & Gould, O. N. (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading*, 31(1), 273–284.
- [23]. McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- [24]. McKeown, M. G. L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford.
- [25]. Miller, A. G., & Gildea, M. P. (1987). How children learn words. *Scientific American*, 257(3), 94–99.
- [26]. Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- [27]. Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: The link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4(1), 151–172. <https://doi.org/10.1515/applirev-2013-0007>
- [28]. MoEYS. (2015a). *Curriculum framework of general education and technical education*. Phnom Penh: MoEYS.
- [29]. MoEYS. (2015b). *Teacher policy action plan*. Phnom Penh: MoEYS.
- [30]. MoEYS. (2017). *Result of grade six student achievement from the national assessment in 2016*. Phnom Penh: MoEYS.
- [31]. Nagy, E. W., & Herman, P. (1984). *Limitations of vocabulary instruction*. Champaign, IL: University of Illinois Center for the Study of Reading.
- [32]. Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82.
- [33]. Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, 44(4), 529–539. <https://doi.org/10.1017/S0261444811000267>
- [34]. Neuman, S. B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in pre-k. *The Reading Teacher*, 62(5), 384–392. <https://doi.org/10.1598/RT.62.5.2>
- [35]. Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- [36]. Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1177/026553220101800103>
- [37]. Seashore, R. (1947). A new light on children's vocabularies. *School and Society*, LXVI, 1(1), 163–164.
- [38]. Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139–152. <https://doi.org/10.1080/09571730802389975>
- [39]. Thorndike, E. L. (1944). *The teacher's word book of 30,000 words*. New York, NY: Teachers College Columbia University.
- [40]. White, T. G., Grave, F. M., & Slater, H. W. (1990). Growth of vocabulary in diverse elementary schools: Decoding and word meaning. *Journal of Educational Psychology*, 82(1), 281–290.
- [41]. World Bank. (2014). *Educating the next generation: Improving the teacher quality in Cambodia*. Washington, DC: World Bank.