International Journal of Linguistics and Translation Studies
Volume 1, Issue 1, 2020
Homepage http://ijlts.org/index.php/ijlts/index

Evaluating the Lexical Load of the Reading Comprehension Texts in EFL Textbooks

Abdelmalek El Morabit

Mohammed V University, Morocco

Lmrabet125@gmail.com

ARTICLE HISTORY

Received: 15/02/2020

Accepted: 26/04/2020

KEYWORDS

Corpus linguistics, frequency, reading comprehension, textbook evaluation, vocabulary profiler.

Abstract

Research has revealed numerous problems regarding the teaching and learning of EFL in Morocco. One of these problems concerns students' performance in reading comprehension tasks. Given the assumption that a large proportion of the problem can be attributed to lexical deficiency, this study sets out to evaluate the vocabulary load of the reading comprehension texts in EFL second-year baccalaureate textbooks. This was actualized through determining how much text coverage students could achieve with their vocabulary knowledge. To this end, a sample of 106 Moroccan bac2 (second-year baccalaureate) students was non-randomly sampled, and their vocabulary knowledge was assessed using an updated version of the Vocabulary Levels Test *(VLT). In parallel, the vocabulary profiles of the reading texts in the three EFL textbooks were described using the Vocabulary Profiler software. Results have shown that students have a rather impoverished vocabulary knowledge of approximately 1317 word families, which have thus revealed a gap between their actual knowledge and the lexical load contained in their textbooks.*

1. INTRODUCTION

Recent research on second language vocabulary acquisition has shown the imperative role vocabulary knowledge plays with regards to boosting students' language skills (Hu, & Nation, 2000) as well as their academic achievement at large (Milton, & Treffers-Daller, 2013). It logically follows from this that a lack of vocabulary knowledge hampers students' performance in different areas of language learning, notably in reading comprehension (Zhang & Anual, 2008).

In line with the last statement, Moroccan students seem to perform poorly in reading comprehension tasks. An anticipated source of this problem may be the lack of the necessary vocabulary knowledge required for effective reading comprehension, perhaps coupled with high lexical requirements from the reading comprehension materials presented in the textbooks. In this way, this study was intended to research the extent to which reading comprehension materials presented in bac 2 textbooks fit students' level regarding vocabulary. To achieve this purpose, students' vocabulary size was tested using a vocabulary levels test. Along with this, an evaluation of the textbooks was carried out by examining the percentages of vocabulary content at each of the first five levels of word frequency and

beyond. Results from both investigations have allowed us to conclude the extent to which reading materials presented in the textbooks correspond to the students' vocabulary knowledge. That said, the study seems to have constructive pedagogical implications for material writers, teachers, learners, and researchers.

2. LITERATURE REVIEW

2.1. Research on vocabulary size

Tang (2007) conducted a study in which he examined the vocabulary level of students at two secondary schools. Similarly, one school was regarded as better than the other, and results show that students in the good school had larger vocabulary sizes. Surprisingly, figures show a larger gap, where the average vocabulary size in one school (3400 words) amounts to almost double the average to the other (1800 words). Of concern to us, however, is only the potential average of the vocabulary size of a high school EFL student, which according to this study ranges from 1800 to 3400 word families.

Canga Alonso (2013) examined the receptive vocabulary size of 92 secondary school Spanish students learning English as a foreign language. Gender was one variable in this study, and the results showed that females had a vocabulary knowledge of 854 of the most frequent 2,000 words in English. Males had a little larger vocabulary size of 1028 words. This estimation should be approached with caution, though, for López-Mezquita's study (as cited in Canga Alonso, 2013) shows a higher vocabulary size range of 941 words for female students in a similar educational context. What this indicates is that, concerning vocabulary coverage required for basic understanding as indicated by Nation (2001), students in both studies would face difficulties in understanding spoken and written texts.

A striking finding insofar as the vocabulary size of EFL high school students is concerned is found in Shin, Chon and Kim's (2011) study which examined the difference between productive and receptive vocabulary knowledge of 402 Korean high school students. What concerns us, though, is the results of the study regarding the receptive vocabulary size of the students. The participants in this study were 11th grade high school learners of around 17 years of age. A bilingual vocabulary size test was administered to test their vocabulary knowledge from the 1st to the 10th 1,000 word bands. Their results demonstrate a vocabulary knowledge of around 6,000 words, which is a relatively elevated score as far as L2 learners' vocabulary size is concerned.

2.2. Vocabulary knowledge and lexical coverage

A study by Nation (2006) has been viewed as the yardstick for measuring text coverage in both listening and reading comprehension. In the study, 98% as a threshold for proper reading comprehension is taken as given, with reference to Hu and Nation (2000). Thus, the study aimed to examine the vocabulary size needed to achieve a 98% coverage in both written and spoken texts. The ultimate results showed that to attain a 98%-word coverage, around 8,000–9,000 word families are needed for reading authentic text, and 6,000–7,000 are required to comprehend a spoken discourse.

Hirsh and Nation (1992) offered different estimations. In their study, they analyzed the vocabulary in three novels to find out how many words are needed to read extensively for pleasure. They concluded that knowledge of 5,000 is a prerequisite for this purpose. Yet another different estimation was offered by Laufer (1992) where 95% coverage was seen to offer a good level of comprehension and was to be achieved with a vocabulary knowledge of 3,000 word families.

Taking the above studies at face value, especially Nation (2006) and Laufer (1992), it seems that there is a huge mismatch between the proposed estimations of the necessary text coverage required for reading comprehension. That is to say, 8,000 words and 3,000 words, as two suggested thresholds for reading comprehension, are by no means close estimations. In reaction to that, Laufer and Ravenhorst-Kalovski (2010) argue for the existence of a

minimal threshold where learners have a vocabulary size of 4,000–5,000 word families, and which results in 95% of text coverage; and an optimal one where learners have a vocabulary knowledge of 8,000 word families allowing the coverage of 98%.

2.3. Research on textbook analysis

Eldridge and Neufeld (2009) is a study that analyzed the vocabulary input contained in a textbook series (SUCCESS series) as well as in a series of graded readers (Oxford Bookworms). The results show that, in both cases, vocabulary is not presented in a way that aligns with recent research findings in the field of vocabulary acquisition and vocabulary pedagogy. For the Success series, for instance, it is found that it comprises only 1,400 of the most frequent 2,000 words in the first four levels collectively. The researchers, therefore, go so far as to recommend doing away with graded readers and coursebooks as sources for reading activities, and they suggest that they be substituted for eReaders, which, as shown also in the study, have "strong potential to help learners develop both an in-depth and productive knowledge of the most frequently used words in English" (p. 224).

Few studies analyzed lexical content in the textbooks with an examination of the vocabulary size of the students. One of these studies is by Teng (2015) who measured 3,105 Chinese students' vocabulary size using the Vocabulary Size Test (Nation & Beglar, 2007) discussed earlier. Results on this test revealed that students had a relatively low vocabulary size of 2,717 word families. Moreover, input from three textbook levels—pre-intermediate, intermediate, and advanced—were analyzed as well using the *Range* program developed by Heatley, Nation, and Coxhead (2002 as cited in Teng, 2015). Results on this program show that the textbooks contained a big deal of low frequency words (42.9%) at the expense of high frequency words. Also, it was shown that only 42% of the words in the textbooks belonged to the first and second frequency levels. That being the case, he also recommends the use of supplementary materials.

3. METHODOLOGY

3.1.Research questions

This study is meant to address the following questions:

- What is the receptive vocabulary size of Moroccan bac2 students?
- What are the vocabulary profiles of reading comprehension texts in bac2 EFL textbooks used in Morocco?
- Does the vocabulary knowledge of students allow them to understand the reading comprehension texts in bac2 EFL textbooks?

3.2.Participants

A total of 124 Moroccan students were non-randomly selected; however, excluded from the analysis were those who did not take the test seriously and those who did not complete the test. Consequently, a total of 106 participants constituted the sample for this study, and they were distributed vis-à-vis their gender and specialty as shown in Table 1.

Table 1. the distribution of the participants across gender and specialty variables (n=106)

		Gender		Total
		Male	Female	
Specialty	Arts	19	33	52
	Sciences	20	34	54
Total		39	67	106

This research attempted to ensure variation in terms of students' specialties in order to offer a general and balanced estimate of the students' vocabulary size.

3.3. Materials

As for textbook evaluation, reading sections from all three nationally used textbooks designed for bac2 students were extracted for analysis. Information about the three textbooks is encapsulated in Table 2.

Table 2. information about the three textbooks analyzed

Textbook	Publisher	Year of publication	Number of pages	Number of units (texts)
Gateway to English 2	Nadia Edition	2007	176	10
Insights into English	Al Massar Edition	2006	175	10
Ticket 2 English	D. I. O EL Hadita	2007	171	10

Four different corpora were assembled where reading comprehension texts were extracted from the three textbooks (see Table 3).

3.4. Research Instruments

3.4.1. The updated Vocabulary Levels Test

The instrument used to measure students' receptive vocabulary size is the updated Vocabulary Levels Test, by Webb, Sasao and Ballance (2017). It is an updated version of the Vocabulary Levels Test (Nation, 1983; Schmitt, Schmitt, & Clapham, 2001) which attempted to remedy some of the limitations that affected earlier versions and to offer a more convenient format. Also, the test came in two equivalent forms (version A and version B) in order to overcome any test retake effects that might result in inflated estimations of students' vocabulary sizes. In this study, we opted for version B of the test.

3.4.2. Vocabulary Profiler

The tool used to analyze the content of the textbook is the Compleat Web Vocabulary Profiler, a computer program developed by Tom Cobb, available at https://www.lextutor.ca/vp/comp/. According to Cobb (2019), Vocabulary Profilers break texts down by word frequencies in a corpus.

3.5.Data Collection Procedures

3.5.1. Vocabulary size testing

The test was administered in four non-randomly sampled high schools to students during the English class sessions. Written instructions were provided in both Arabic and English as well as information about the nature of the test and the purpose behind it. The participants were recruited by asking language teachers to have their students take the test during the last 40 minutes of their classes. The tests were handed to the teachers just before the start of the session and were obtained back immediately after the session.

3.5.2. Corpus creation

Concerning the analysis of the textbooks' vocabulary profiles, four small corpora were compiled. Reading comprehension texts were extracted from the three textbooks and were assembled into three Plain Text files named after the textbooks they were taken from, so the three corpora were named *Ticket*, *Insights*, and *Gateway*. These three corpora were then assembled into one single compact file to represent reading comprehension texts at the baccalaureate level at large, which constituted the fourth corpus named *Compact TIG*. Information about each corpus is given in Table 3.

Table 3. information about the four corpora

Name of the corpus	Source	Number of running words (tokens)
Ticket	Ticket 2 English (2007)	4,718
Insights	Insights into English (2006)	5,709
Gateway	Gateway to English 2 (2007)	4,832
Compact TIG	All three of the above	15,258

The four corpora were uploaded successively to the Lextutor Vocabulary Profiler for analysis.

4. RESULTS

4.1. Receptive vocabulary knowledge

In connection with the first research question, this section provides descriptive statistics of the receptive vocabulary size of the students taking part in the study. The students' overall estimated vocabulary size is presented in Table 4.

Table 4. descriptive statistics for Students' receptive vocabulary size

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Students' receptive vocabulary size	106	3067	367	3433	1317.30	679.842
Valid N (listwise)	106					

As shown in Table 4, the mean score (1317.3) represents the average receptive vocabulary size of the students (N=106). Based on this result, it can be said that students have a receptive vocabulary size of around 1317 word families in total. However, there seems to be a large variability among the scores obtained by the students on the receptive vocabulary knowledge test in that the standard deviation (SD=679.842) shows a considerable deviation of scores from the mean. This variability can also be observed very clearly in the big difference between the minimum (367) and the maximum scores (3433) obtained by students. Considering that the instrument used measures students' vocabulary in relation to the most frequent 5,000 word families, the scores obtained reflect a fairly low vocabulary profile on the part of the students, where the maximum score does not even reach 3,500 word families in total.

In order to further investigate the variability among the scores of the students, frequencies of scores seem worth looking at. Figure 1 gives a general overall idea of the shape of the distribution of scores.

Figure 1 the distribution of the students' vocabulary size scores

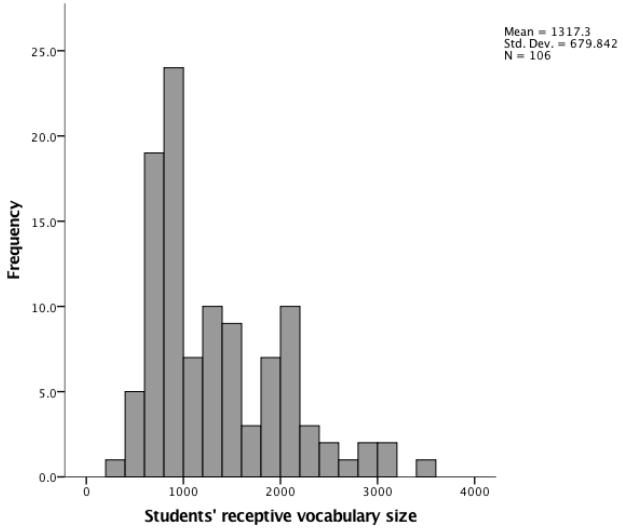


Figure 1 appears to give an elegant image of the distribution of scores. First, most of the scores seem to be centered in the first two thousand, with a peak emerging in the area of 700-1,000, which indicates that most of the scores are in fact centered in this particular area. In addition, there seem to be some relatively high scores which, according to the figure, appear in the area of 3,400-3,500. These are most likely outliers, exhibiting extreme scores relatively to other scores.

4.2.Students' receptive vocabulary knowledge at K1-K5 levels

Bearing in mind the general purpose of the study, it appears that students' scores concerning different levels of word frequency are of more concern. Thus, students' scores will be presented collectively in terms of K1-K5 levels in order to set the ground for further analysis in the light of textbooks' content. Table 5 is presented to demonstrate students' scores at all five levels of word frequency.

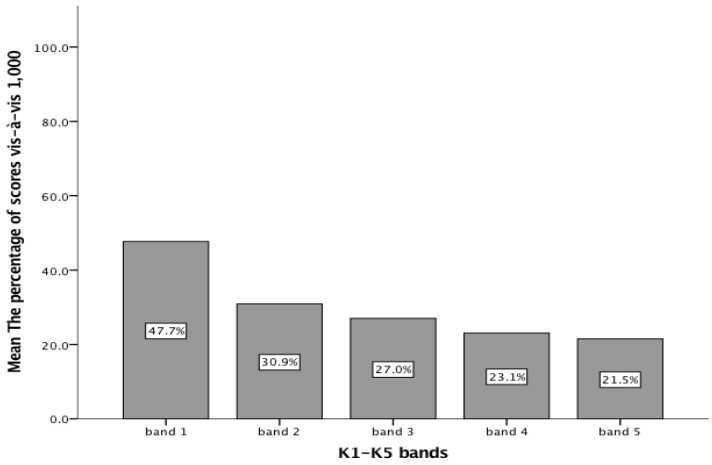
Table 5 descriptive statistics for the students' scores at K1-K5 levels

	N	Minimum	Maximum	Mean	Std. Deviation
Band 1	106	67	1000	477.04	276.920
Band 2	106	33	1000	308.81	239.457
Band 3	106	33	833	269.81	175.874
Band 4	106	33	867	231.45	173.225
Band 5	106	0	733	215.41	168.603
Valid N (listwise)	106				

As shown in the table, students' scores at all levels seem to be considerably low, considering that each frequency band consists of 1,000 word families in total. As for level 1, which corresponds to the most frequent 1,000 word families in English as represented in the BNC-COCA word list, the mean score shows that students have a knowledge of only 477 word families out of 1,000 most frequent words. This shows a significant deficiency in their basic word knowledge. The minimum (67) and maximum (1,000) scores, along with the standard deviation (SD=276.9), denote a considerable dispersion of the scores at this level, and hence a large degree of variation in the word knowledge among students.

As mentioned earlier, the mastery of a word frequency level requires students to score 97% at that level, which amounts to knowledge of around 966 word families. That said, Figure 2 shows the percentages of students' scores at 1K-5K levels.

Figure 2 percentages of students' scores at each frequency level



From the figures, we conclude that students' scores are far below the required percentage for the mastery of different word levels. At level 1, for instance, students score 47.7%, which is far below the percentage of 97% required for the mastery of the level. Similarly, scores at all other levels revolve around the area of 20-30% of word knowledge, which indicates utter lexical deficiency at those levels, and very scant familiarity with the most frequent 5,000 words necessary to function in English.

4.3.Description of the vocabulary profiles of the four corpora

In this section, an analysis of each of the four corpora is provided at a time so that we can observe if there is a difference between the three textbooks. It is also important to reiterate that the program used for analysis distributes the text used at all 25 word levels of the BNC-COCA-25; yet, for purposes of the study, levels beyond 5,000 will be included in one row indicating words of lower frequency.

4.3.1. Analysis of Ticket

This corpus concerns the analysis of the reading materials from *Ticket 2 English* (Hammani, et al. 2007). Table 6 shows the frequency of words in the Ticket corpus across the 25 frequency bands in the BNC-COCA.

Table 6 word frequency in Ticket

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1 Words :	497 (45.2)	727 (48.79)	3555 (75.3)	75.3
K-2 Words :	243 (22.1)	349 (23.42)	584 (12.4)	87.7
K-3 Words :	234 (21.3)	268 (17.99)	397 (8.4)	96.1
K-4 Words :	45 (4.1)	49 (3.29)	61 (1.3)	97.4
K-5 Words :	40 (3.6)	41 (2.75)	42 (0.9)	98.3
K-6 – K-25 Words :	40 (3.7)	40 (0.74)	57 (1.2)	99.4
Off-List:	??	21 (1.41)	22 (0.47)	99.87
Total (unrounded)	1099+?	1490 (100)	4718 (100)	≈100.00

As the table shows, 497 word families belong to the first one thousand frequency band, which accounts for 45.2% of the word families in the corpus. This amounts to almost half the total word families in the rest of the frequency bands in the corpus. This is typical and can be accounted for by the fact that most function words, which by nature occur in large frequencies, belong to that level. Also, the subsequent two frequency levels, namely K2 and K3, contribute by a considerable amount to the total of word families in the corpus. They make up 22.1% and 21.3% of the total word families respectively. Looking at tokens, these three first levels together make up 96.1% of the total of tokens in the corpus, which amounts

to saying that a learner with a mastery of the first three levels of word frequency will achieve 96.1% of token coverage of the texts included in the corpus.

Generally, for Ticket, it has been shown that a fairly good percentage of the text can be found in the first frequency band. This text percentage drops drastically as we move towards lower frequency levels, which is quite normal. Around 3% of the remaining text is distributed among the 4K-25K levels.

4.3.2. Analysis of Insights

The second corpus, entitled *Insights*, is made out of texts from the textbook *Insights into English* (Najbi, et al. 2006). Here, again, the text will be viewed in terms of different frequency bands, and the amount of text at each of these bands will be considered. Initially, Table 7 shows word frequency in the corpus in relation to the K1-K25 levels in the BNC/COCA.

Table 7 word frequency in Insights

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1 Words :	505 (44.3)	766 (47.93)	4305 (75.4)	75.4
K-2 Words :	262 (23.0)	381 (23.84)	708 (12.4)	87.8
K-3 Words :	231 (20.3)	280 (17.52)	466 (8.2)	96.0
K-4 Words :	49 (4.3)	54 (3.38)	85 (1.5)	97.5
K-5 Words :	30 (2.6)	34 (2.13)	51 (0.9)	98.4
K-6 – K-25 Words :	63 (5.5)	64 (4.0)	72 (1.26)	99.5
Off-List:		22 (1.38)	22 (0.39)	99.89
Total (unrounded)	1140+?	1598 (100)	5709 (100)	≈100.00

Very similar to the previous analysis, *Insights* exhibits a moderate vocabulary profile concerning the first frequency band in so far as the total word families are concerned. Moreover, the second and third levels combined add up to almost the total of the first level. Levels K4-K25 now occupy almost 12.4% of the text. Regarding the frequency of tokens, differing results are put forth. K1 presents 4305 word tokens (75.4%), while the second and third levels present 708 (12.4%) and 466 (8.2%) tokens respectively. These give clearer estimations of how much coverage can be achieved by students with different vocabulary sizes. The off-list row indicates that a very small amount of words is missing from the lists. Examples of words classified as off-list in this corpus are *online*, *schematized*, and *jokester*.

4.3.3. Analysis of Gateway

The third corpus analyzed in this study concerns texts from the textbook *Gateway to English 2* (Hassim, et al. 2007). We will proceed by looking at the vocabulary profile of the corpus in terms of word frequency levels, in addition to the other related indices. To start with, Table 8 demonstrates word frequencies in the corpus.

Table 8 word frequency in Gateway

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1 Words :	453 (45.5)	675 (47.97)	3569 (73.9)	73.9
K-2 Words :	225 (22.6)	330 (23.45)	644 (13.3)	87.2
K-3 Words :	225 (22.6)	278 (19.76)	452 (9.4)	96.6
K-4 Words :	40 (4.0)	42 (2.99)	55 (1.1)	97.7
K-5 Words :	19 (1.9)	21 (1.49)	26 (0.5)	98.2
K-6 – K-25 Words :	33 (3.3)	35 (2,48)	42 (0.9)	98.9
Off-List:		42 (2.99)	44 (0.91)	99.81
Total (unrounded)	995+?	1407 (100)	4832 (100)	≈100.00

Unsurprisingly, the results from the analysis of this corpus do not differ widely from the analysis of the previous two corpora. At the K1 level, 453 family words are found, amounting to 3569 word tokens. Since a word family is instantiated by several tokens, and counting word families in a corpus does not take account of repeated occurrences, the percentages of the families and tokens coverage of the corpus differ (45.5% word families and 73.9% word tokens). Put differently, learners with full mastery of K1 level will achieve 45.5% of word family coverage in this corpus and of 73.9% of tokens. Moreover, a mastery of 453 word families in the corpus will also guarantee coverage of all 3569 tokens in the corpus.

4.3.4. Analysis of Compact TIG

The most important of the corpora designed for this study is entitled *Compact TIG*. It comprises texts from all three textbooks designed for bac2 level, and it is a compilation of all the three corpora analyzed previously. It is meant to be representative of all reading comprehension texts, and hence of the targeted lexis, designed for bac2 students. Exploring word frequency in Compact TIG will offer insights into the suitability of the lexis to the students' capacities. Table 9 below offers a summary of the word frequency at the different levels of the BNC-COCA word list.

Table 9 word frequency in Compact TIG

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1 Words :	686 (35.5)	1239 (40.12)	11428 (74.9)	74.9
K-2 Words :	437 (22.6)	753 (24.38)	1936 (12.7)	87.6
K-3 Words :	482 (24.9)	675 (21.86)	1315 (8.6)	96.2
K-4 Words :	120 (6.2)	133 (4.31)	201 (1.3)	97.5
K-5 Words :	77 (4.0)	85 (2.75)	119 (0.8)	98.3
K-6 – K-25 Words :	131 (6.8)	136 (4.36)	171 (0.3)	99.2
Off-List:		??	88 (0.58)	99.78
Total (unrounded)	1933+?	3088 (100)	15258 (100)	≈100.00

From the table, it is shown that 11428 (74.9%) out of the 15258 tokens in the corpus belong to the first frequency band. An additional 12.7% of the tokens occurs at the second frequency band and 8.6% at the third frequency band. In total, these three levels, constituting the high frequency vocabulary, account for 96.2% of the vocabulary contained in the corpus. This indicates that a somewhat high proportion of the text is centered at these levels.

5. DISCUSSION

Looking at the scores obtained in this study in more detail, we find that the above conclusions apply not only to the overall score but also to individual frequency levels. Students' scores at all five frequency bands are particularly low. However, if we consider the descent of the cline across the five frequency levels (see Figure 2), there appears to be an abnormal closeness of the scores across the five different levels. This can be attributed to a variety of reasons. First, we might assume that the vocabulary in the syllabi used in public high schools does not take into account the word frequency as a major criterion for selecting and sequencing vocabulary. It is also possible that vocabulary instruction is characterized by randomness and is not designed in a way that responds to the students' actual needs. Lastly, it is also legitimate to assume that the issue lies with students and the way they handle their vocabulary acquisition.

A legitimate question at this juncture would be whether students definitively have no mastery of any of the five levels investigated. Revisiting the threshold for level mastery proposed by Webb et al. (2017)—97%, students have been reported not to have a mastery of any of the frequency levels. Still, there are differing opinions in the literature concerning this threshold. In their revised version of the VLT, Schmitt, Schmitt, & Clapham (2001) suggested that a threshold for the mastery of a level would be at 87%. This amounts to saying that knowledge

of a minimum of 866 words out of 1,000 would indicate mastery of that level. In 2003, however, in personal communication with Xing and Fulcher (2007), Schmitt suggested 80% as a cutting point for level mastery. Despite this low suggested threshold, the participants in this study, with the highest percentage of vocabulary knowledge attained at the first level amounting to 47%, seem away below the mastery of any of the levels. Even when compared to the lowest percentage of level mastery proposed in the literature by Nation (1983), that is 66%, students seem far from any level mastery.

Perhaps the best way to understand the findings of our study, and thus see how low the vocabulary size of students is would be to compare them to other studies conducted in a variety of contexts. Concerning the overall vocabulary size of students, participants in this study exhibited a lower vocabulary size than those belonging to the "weak" school in Tang's (2007) study. Those participants were secondary school students in China where Chinese was used as a medium of instruction and that a small number of students from those schools got admitted to universities. What this implies is that, considering the comparatively lower vocabulary size of Moroccan students, the high schools sampled for this study would legitimately be considered 'weaker' schools. The same conclusions are drawn from contrasting our results with Shin, Chon and Kim's (2011). The participants in their study were in 11th grade of high school and of somewhat a similar age as the participant in our study. These students exhibited knowledge of around 6,000 word families. This simply pinpoints how impoverished Moroccan high school students' vocabulary knowledge is and, by entailment, calls for change to be made. That stated, the importance of this study becomes more palpable.

We have previously seen different percentages of text coverage required for text comprehension. In general, we have stated that a rule of thumb is 95-98% coverage of text is required for reading comprehension. In the case of this study, achieving 98% coverage required students to have a mastery of all the first five frequency levels in the BNC/COCA. This was not the case. But, considering that 98% coverage will allow unassisted reading, a lower percentage can be used considering that the bac2 students can have assistance from the teacher, among other potential sources. In this case, Laufer (1992) suggested that 95% coverage can allow students to comprehend with a little need for assistance. In order to reach this percentage of text coverage in the Compact TIG corpus, students are required to have a mastery of the first three frequency bands. Considering table 5, again, none of the students participating in the study seems to have a mastery of all three levels. Therefore, none of the students has a vocabulary knowledge needed for 95% coverage, which is considered by Laufer and Ravenhorst-Kalovski (2010) as a minimal threshold for text comprehension. Considering the same table, very few students showed a mastery of the first two levels, which barely allow for 87% text coverage; and this, according to Nation (2001), can only provide a minimum degree of comprehension or no comprehension at all.

On a final note, the findings of this study seem to shed light on very important aspects of TEFL in Morocco, namely students' acquisition of vocabulary and the use of corpus tools to evaluate reading materials. The results generally show that students are not yet ready to handle materials that require knowledge of less frequent vocabulary. Therefore, a few implications are suggested in the subsequent section.

6. CONCLUSIONS AND PEDAGOGICAL IMPLICATIONS

This study generally aimed at investigating whether or not the vocabulary in reading comprehension texts in bac2 textbooks is suitable for the learners. For this purpose, students' vocabulary knowledge was measured using Webb et al.'s test (2017). The test allowed us to investigate the students' overall vocabulary size as well as their knowledge at different frequency levels. Additionally, four corpora were created in order to analyze the vocabulary profile of the three bac2 textbooks. We ultimately found out that there is a large gap between the students' vocabulary knowledge and the vocabulary contained in the reading

comprehension sections in the textbooks. This study can also be seen as an attempt to draw attention to a very significant area of research in SLA, which is vocabulary research, that can offer very interesting implications for teaching English as a foreign language. Regarding the methodology, this study has hopefully made clear the utility of corpus linguistics in ELT as an area that offers very powerful research tools that can be exploited for a variety of purposes.

The findings of the study seem to have pedagogical implications for learners, teachers, and material designers. First, students are encouraged to invest more effort in learning new words. And because there is no optimal technique for learning vocabulary, learners should explore different techniques and see which ones work best for them. Equally important, teachers are urged to use supplementary materials in their classes in order to boost students' vocabulary knowledge. Also, teachers are urged to evaluate the materials to be used in class and make changes if the materials turn out to be unsuitable. Finally, material designers should take this into account when designing materials for students. Research findings on vocabulary should be considered when designing different syllabuses, and vocabulary incorporated in these syllabuses should be closely monitored and controlled. Moreover, the frequency of lexical items should constitute one of the criteria when selecting and grading vocabulary items in textbooks.

REFERENCES

Canga Alonso, A. (2013). Receptive vocabulary size of secondary Spanish EFL learners. *Revista de Lingüística y Lenguas Aplicadas, 8,* 67-75. doi:10.4995/rlyla.2013.1180.

Cobb, T. (2019) VocabProfilers [computer program]. Retrieved from https://www.lextutor.ca/vp/

Eldridge, J., & Neufeld, S. (2009). The graded reader is dead, long live the electronic reader. *The Reading Matrix, 9* (2), 224-244.

Hammani, M., Ahssen, S., & Tansaoui, L. (2007). *Ticket 2 English. Casablanca*. Dar Ihiae Oloum EL Hadita.

Hassim, M., Blibil, M., & Rasmy, A. (2007). *Gateway to English 2. Rabat*: Nadia Edition.

Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? Reading in a Foreign Language, 8, 689-689.

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

Laufer, B. & Ravenhorst-Kalovski (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language 22(1).* 15-30. doi:10.5353/th_b3195924

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.

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review, 4*, 151–172. doi:10.1515/applirev-2013-0007

Najbi, M., & El Haddad, M. (2006). *Insights into English.* Rabat: Al Massar Edition.

Nation, I. S. P. (1983). *Testing and teaching vocabulary*. Guidelines, 5, 12-25.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-82. doi:10.3138/cmlr.63.1.59.

Nation, I. S. P., & Beglar, D. (2007). *A vocabulary size test. The Language Teacher*, 31(7), 9–13.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. Language Testing, 18, 55–89. doi:10.1177/026553220101800103

Shin, D., Chon, Y. V., & Kim, H. (2011). Receptive and productive vocabulary sizes of high school learners: What next for the basic word list? English Teaching, 66 (3), 127-152. doi:10.15858/engtea.66.3.201109.123

Tang, E. (2007). An Exploratory study of the English vocabulary size of Hong Kong primary and junior secondary school students. *The Journal of Asia TEFL, 4*(1), 125-144.

Teng, F. (2015). An evaluation of EFL students' vocabulary size and their textbooks: A case study of vocational college students in China. Humanizing Language Teaching, 17 (5), 26-41.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics, 168*(1), 34-70. doi:10.1075/itl.168.1.02web

Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. System, 35(2), 182–191. doi:10.1016/j.system.2006.12.009

Zhang, L. J., & Anual, S. B. (2008). The role of vocabulary in reading comprehension: The case of secondary school students learning English in Singapore. *RELC Journal, 39*(1), 51-76, doi:10.1177/0033688208091140

Author Biodata

Abdelmalek El Morabit ***is a Ph.D. candidate at the faculty of letters and humanities Mohammed V, Rabat. He is also a TESOL-certified English teacher, and he has taught both English and Arabic as foreign languages at different institutions. He is mainly interested in second language acquisition and corpus linguistics, and he is currently researching the developmental sequences in second language vocabulary and vocabulary network building.***