Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327
مجلة اللغة الانكليزية في العالم العربي

Arab World English Journal (AWEJ) Volume 9. Number 2. June 2018 Pp. 234 -250
DOI: https://dx.doi.org/10.24093/awej/vol9no2.16

Lexical Frequency Effect on Reading Comprehension and Recall

Nadia Nouri
Faculty of Education, Mohammed V University
Rabat, Morocco

Badia Zerhouni
Faculty of Education, Mohammed V University
Rabat, Morocco

Abstract

The present study investigates the effect of lexical difficulty, as measured by frequency, on reading comprehension and recall. It also estimates the relationship between vocabulary size, vocabulary depth, reading comprehension, and recall. To this end, 80 English as a foreign language (EFL) undergraduate university students are administered three standardized instruments including two vocabulary tests and a reading comprehension test. The latter comprises two similar passages (in terms of length and topic) one of which is adjusted by replacing 18% of its words by their low-frequent synonyms, and both passages are estimated lexically to measure their difficulty. Paired samples t-test results show that comprehension and recall are significantly low (p<.001) in the modified passage. This finding further confirms that lexical frequency measure is an effective estimate in determining reading material difficulty. Furthermore, for the second objective, Pearson product-moment analysis reveals a significantly high correlation between size and comprehension, a moderate to low correlation between depth and comprehension, and a moderate correlation between depth and recall tasks. Consequently, the study sugguests estimating the complexity of EFL reading academic material with lexical difficulty measure using frequency criterion to cope with students' reading deficiencies, and encourage explicit lexical instructions at EFL undergraduate university levels mainly.

Keywords: lexical frequency, reading comprehension, recall, vocabulary depth, vocabulary size

Cite as: Nouri,N., & Zerhouni, B. (2018). Lexical Frequency Effect on Reading Comprehension and Recall. *Arab World English Journal, 9*(2). DOI: https://dx.doi.org/10.24093/awej/vol9no2.16

Introduction

Research has shown that lexical knowledge constitutes a "pressure point" in reading comprehension (i.e. the linkage between word recognition and text comprehension) (Perfetti, 1985; Perfetti & Hart, 2002; Perfetti, 2007). This is the reason why several researchers have been interested in estimating text difficulty for comprehension. They have designed traditional readability formulas (e.g. Lively & Pressey, 1923; Patty & Painter, 1931; Ojemann, 1934) as well as modern computerized measures (e.g. Lexile, 2007; Milone & Biemiller, 2014; McNamara et al., 2014). These assessment tools show high reliability with reading complexity (Nelson et al., 2012). However, given the different features estimated in these formulas such as word syllables, sentence length, and frequency, it is unclear which aspect better determines reading complexity. An alternative to these formulas is testing text difficulty by word complexity estimate (Chen & Meurers, 2016, 2017). This research paper has opted for this option because it seems more appropriate to the study objectives. The latter, therefore, focuses on investigating the extent to which word difficulty, as measured by frequency, affects reading comprehension and recall, and examining the relationship between Moroccan EFL students' vocabulary size, depth, reading comprehension, and recall.

Literature Review

Vocabulary Knowledge: Aspects Definition and Relationship

Some researchers contend that knowing a word is characterized by its form, meaning, and use (e.g., Alderson, 2000; Read, 2000; Schmitt, 2000; Nation, 2001). For other researchers, word knowledge includes meaning, usage, grammatical and semantic constraints, associations, and morphology (e.g., Koda, 2005; Bernhardt, 2005; Zhang, 2012). These frameworks are all appropriate to define the construct of interest. However, a better understanding of a word relies on dissecting lexical knowledge into small manageable aspects for insightful testing. In this study, vocabulary knowledge is conceptualized by size and depth dimensions. Vocabulary size is defined as knowing the single meaning of a word, and is estimated by Vocabulary Size Test (VST) (Nation & Beglar, 2007). Vocabulary depth, on the other hand, refers to the degree in which a word is well known, and is measured by Word Associates Test (WAT) (Read, 1993). Anderson and Freebody (1981) are the pioneers of this framework. They advance that internalizing words begins with a shallow knowledge, and as learners proceed in learning, their word mastery deepens as well. Thus, size and depth seem to constitute a continuum, even though some researchers tend to contrast them as two distinct dimensions. In the following paragraphs, discussion will be centered on the relationship between these two aspects of vocabulary knowledge as tested by Vocabulary Levels Test (VLT) and Word Associates Test (WAT) in particular.

One of the studies that investigated this issue was conducted by Qian (1998) who investigated the relationship between vocabulary depth, size, and reading comprehension. The sample included 74 learners attending English as second language (ESL) intensive classes in Ontario Universities. His results revealed a significantly strong correlation between VLT and WAT r=.82, p<.05. After some modifications on the depth measure, Qian concluded that depth adds a unique contribution to reading comprehension. One probable reason for such a finding is that some words from the reading test are also part of the depth measure. Therefore, participants

who responded correctly to the target words in the depth test might probably have found it easier to comprehend these items in the reading test context. To account for this side effect, Qian varied the order of the tests' administration in each testing session by administering, for example, the reading test before depth and size tests in one session and administering depth test first followed by size and reading measures in another session. However, possible learning of the target words may still happen if depth test is administered before reading.

In a sequel study, Qian (2002) investigated a large sample composed of 217 students attending an intensive ESL program at Toronto University and coming from nineteen first language (L1) backgrounds. Results revealed that correlation between VLT and WAT was not as high as in his 1998 findings, with a coefficient of r=.70, p< .01. One possible explanation of such finding might be the diversity of participants' L1 backgrounds. It is quite probable that one sub-sample of participants might have a better developed lexical knowledge than other sub-samples given different learning experiences and the amount of lexical input they might have accumulated. Besides, lower correlations have been found as well in the VLT-WAT comparison conducted by Zhang and Horiba's studies (2012) on Chinese learners with a coefficient around r=.52, p<.01.

It should be noted, based on the findings of most studies conducted on this issue, including those reported in this section that no negative correlation has been revealed between vocabulary size and depth. This may indicate that these two lexical dimensions are not separate entities as supported by a number of researchers (Read, 2004; Milton, 2009; Henriksen, 1999). However, size and depth seem to converge more largely at advanced proficiency levels.

Concerning measurement of lexical size, there is a fundamental difference between a diagnostic measure and an overall performance test. Thus VLT, despite its high reliability and close relation to depth is only a diagnostic test, as it merely estimates 'slices' of lexical items across five frequency levels (i.e. 2000, 3000, 5000, the university word list, and 10000 word level). Therefore, even though it is widely used in lexical studies, it should not be interpreted as an estimate or even a proxy measure of size (Nation, personal communication). For this reason, the present study has opted for an alternative estimate: Vocabulary Size Test (VST) (Nation & Beglar, 2007). It consists of 14 frequency levels, including 10 items per level. However, given the limited lexical competency of the present study participants, and based on the test piloting, it was judged adequate not to include the last four levels in the administered version though Nation (2012) recommends completing the whole test. This omission was based on the researchers' belief that it would not distort the overall measure, and, at the same time, would reveal the actual lexical proficiency of the participants. Another reason for choosing VST is its ability to distinguish different vocabulary sizes, which plays a highly important role in reading comprehension investigation (Stuart & Kramer, 2015).

Few researchers working on vocabulary size and depth have been concerned with text difficulty even though performance in reading comprehension is not only related to readers' lexical knowledge, but also to the complexity of the reading material. Generally, the latter issue is tackled through deletion, instruction, and substitution studies of vocabulary. However, to the best of the

researchers' knowledge, it has not been fairly related to vocabulary size and depth. The following sections discuss the role of lexical difficulty in estimating text complexity, and shed light on word information level of texts, which is central to the present study.

Lexical Frequency, Word Information Level, and Reading Comprehension

Based on the 'Lexical Quality Hypothesis' (Perfetti & Hart, 2002), word frequency is the result of the benefits of language learning experience. Higher frequency words make their meaning and form "a more stable constituent compared with the lower frequency meaning" (Perfetti et al., 2002, p. 194). This concept is associated with the semantic aspect of lexis mainly. Besides, Anderson and Freebody (1979) define frequency as "a characteristic of a word which probably is very strongly related to the chances that the word will be known" (p. 23). For this reason, performance in reading comprehension can be analyzed from a lexical perspective.

Frequency researchers have conducted several studies to determine the cut-off line between high and low-frequency levels, using English corpora. For example, Nation (2001) suggests that 2000-word families are the upper line of high frequency. He also contends that a critical goal of 8000-word families is what learners need to read complex written texts (Nation, 2006). Academic material is specifically the most demanding in reading as it may require this threshold to attain comprehension and learning (Schmitt, 2010). A good question to ask here is whether Moroccan EFL learners reach this threshold at the undergraduate university level. Another point worth inquiring in further studies is whether, with this objective, a cut-off line of high frequency is still 2000-word families. Schmitt (2010) indicates that it is difficult to stop at this category line of high frequency, and suggests three frequency bands to bridge the gap between high and low frequencies: 1000-3000 word families range is high, 3000-7000 is mid, and 8000-10000 is low frequency (Schmitt, 2014). Based on Schmitt (2014) categorization, academic EFL material may fall within the mid-low category.

Furthermore, text coverage is highly related to comprehension. Thus, in reading, 95% (Laufer, 1989) and 98% (Hu & Nation, 2000) of the running words in texts should be familiar to readers. In this regard, Schmitt (2010) clarifies that these rates have become a rule of thumb thanks to the number of studies reporting them and that specialists of texts coverage did not intend in the first place to make these figures a reference given the limited number of participants in their studies and the small number of texts analyzed. One should question, therefore, whether these ratios are universal for reading comprehension. Practically, it seems unrealistic to teach systematically the lexical knowledge needed for 95% or 98% of text coverage in an EFL context.

Mezynski (1983), on the other hand, points out interestingly that the proportion of unknown words within a text is not as important as that of words with high information level. This claim puts coverage equations into query. From these conflicting perspectives, one may question whether readers need to know only the words with high information level to achieve comprehension. Unlike marginal words, items with high information value give significant information about the meaning of sentences in a passage. Adams and Bruce (1980) clarify this by the following example "the discovery of a number of fossilized porbeagles in Kansas is intriguing" (p.8). If one does not know

that porbeagles are 'large ocean-dwelling fish' and that 'intriguing' means 'fascinating', it will be difficult to comprehend the meaning of this sentence (Mezynski, 1983, p. 261). Though this seems self-evident, it is still unclear whether such category of words serves comprehension or percentages of text coverage mentioned before do so.

Researchers using cloze technique reveal the importance of lexical information value. They have shown that skilled readers are able to understand texts where up to 30% of the words have been omitted (Nacke, 1970; Martin & Pantalion, 1973). In this regard, Mezynski (1983) states that "the important factor in these deletion studies is the information level of the words. The absence of low information words does not hinder comprehension considerably" (p. 261). Similarly, Stratton and Nacke (1974) indicate that "readers more frequently encounter problems with inadequate vocabulary when a word of high information value is unfamiliar" (p.186 in Mezynski, 1983, p. 261).

To examine the effect of word frequency on reading comprehension, Marks, Doctorow, and Wittrock (1974) conducted a reading experiment wherein 15% of lexical items in five texts were substituted into low and high frequency levels. T-test analysis results showed that when learners were exposed to 'hard' versions of the texts followed by 'easy' ones, comprehension was remarkably low at the beginning but improved later. However, when 'easy' versions were given first, there was no difference in comprehension. In this study, one would draw two conclusions: learners seem to learn low-frequency words meaning from their high frequent synonyms when easy versions of texts are administered first, and low-frequency words seem to deteriorate comprehension performance when no prior simplification was provided.

Anderson and Freebody (1981) reported two experiments of reading using sixth-grade pupils. The first experiment controlled the number of unknown words, and the second one varied the importance level of unfamiliar items with regard to the passage. The post-reading tasks required were free recall, writing a summary statement, and sentence recognition (i.e. deciding whether some ideas belong to the texts). Findings indicated in the first experiment that comprehension was not highly affected when only one out of six words (1/6) was replaced with a low-frequency synonym. However, when one out of three (1/3) words was substituted, comprehension declined significantly in the sentence recognition task especially. This means that in 1/3 there are two chances of encountering an unfamiliar word in a six-word string. In other words, the density of unfamiliar words influences comprehension. One question to raise concerning these experiments is whether the substituted words are key items in the texts. The variance of word information value is another crucial point to highlight in the second experiment. The effect of unfamiliar words with high information level is unclear since they are altered by low information ones. Thus, the variety of these interactions may cause some confusion in the results. For this reason, interpretations of the second findings should be cautious.

Measuring Text Difficulty with Lexical Complexity

As mentioned earlier, a large number of readability formulas have been designed to estimate text complexity. Several criteria have been employed in these measures causing different

interactions between aspects. The latter makes it difficult to know which element determines text complexity most. None of the formulas, however, focuses exclusively on the effect of frequency on reading comprehension (Chen & Meurers, 2016).

There is a consensus among specialists in reading research that lexical coverage and vocabulary knowledge are strong predictors of reading comprehension (Laufer, 1992; Qian, 1998, 2002; Nation, 2001, 2006). A reader's vocabulary knowledge is highly attributed to the amount of lexical input to which they have been exposed (Chen et al., 2016). This is referred to as 'frequency effect' (Perfetti et al., 2002; Chen et al., 2016, 2017). Ryder and Slater (1988) have confirmed that frequency predicts word difficulty. It seems thus that high-frequency words are perceived more quickly than low frequency ones (Rayner & Duffy, 1986). Therefore, estimating text difficulty on a frequency basis is fundamental in reading research.

Frequency effect is discussed in the literature based on a cognitive model. It assumes 'high activation' for frequent words in the mental lexicon while a reader is processing written texts (Just & Carpenter, 1980). This goes in line with the 'Lexical Quality Hypothesis' (LQH) initially labeled as 'Verbal Efficiency Theory' (Perfetti, 1985) and later developed by Perfetti and Hart (2002). LQH suggests that words with high representation are likely to be 'effortlessly' accessed by readers. This means that a word form and meaning are strongly linked to one another and can be retrieved whenever the item is encountered in reading. Moving beyond this basic notion, Ellis (2012) states "frequency distribution of input is a key determinant of language acquisition, with regularities emerging through the learner's exposure to the distributional characteristics of the language input" (p.85). Hence, estimating text complexity in terms of word difficulty provides insightful considerations to EFL acquisition.

While findings reviewed here suggest the crucial role of low-frequency words in reading comprehension, it is unclear what type of information level these words have. This explains the choice of the present study to substitute high information value words by low-frequent synonyms. Furthermore, estimating the whole text difficulty with frequency measure highlights the role of high information value words in comprehension, and puts into question text coverage of 95 and 98% suggested by Laufer (1989) and Hu & Nation (2000) respectively as mentioned earlier.

The Current Study

As mentioned earlier, the present study aims to investigate the relationship between lexical frequency, reading comprehension, and recall. Approximately, 18% of the most important words in one of two texts used in the study have been substituted by low frequent synonyms to examine the extent to which a small proportion of unknown words may influence participants' reading comprehension (Mezynski, 1983) and recall (Anderson & Freebody, 1981). Text difficulty is thus measured in terms of word frequency to highlight the importance of lexical frequency model in estimating readability at EFL university level (Chen & Meurers, 2016). Frequency here is mainly characterized by the semantic variable, referring to readers' ability to relate text words to their pre-existing vocabulary knowledge (Marks et al., 1974; Ryder & Slater, 1988). Semantic variable of word frequency has been found to account for the variances in reading comprehension (Marks et

al., 1974). Also, based on the Lexical Quality Hypothesis (Perfetti & Hart, 2002), this investigation examines the relationship between vocabulary size, depth, and reading comprehension.

Methods

Participants

Eighty EFL third-year university students from Mohammed V University-Rabat, Morocco participated in this study. The sample consisted of 47 females (58.75%) and 33 males (41.25%) with an age range between 20 and 27 years.

One single university level has been involved in this study for two reasons. First and foremost, third-year students are assumed to constitute the ideal target population for the present research given the considerable amount of lexical knowledge they have accumulated thanks to the language input they have been exposed to during their EFL learning experience. Second, because of the study complexity, it has been judged preferable to limit its scope to one educational level.

Measures

Data has been collected based on two lexical tests and one reading comprehension test. Each of these instruments is discussed in the following sections.

Vocabulary Size Test (VST) designed by Nation and Beglar (2007) is used to examine students' vocabulary size. It has a high Cronbach's alpha reliability of 0.92, based on the 10th 1000-word families used. As mentioned before, the last four levels were removed due to their difficulty level for the study sample. Besides, VST estimates mainly acquired vocabulary. In each item, the test provides non-contextualized sentences and closely related options in meaning. Thus, informed guessing is difficult to make (Nation, 2012). In scoring, 1 point is worth every correct answer and the total score is multiplied by 100.

Word Associates Test (WAT) developed by Read (1993) is used to measure students' vocabulary depth. It examines how well learners know vocabulary in three word aspects: synonymy, polysemy, and collocations. Forty adjectives are tested in this measure with an indirect insight into nouns. In scoring, each correct response is worth 1 point with originally a 160 maximum possible score. After piloting the test, however, four difficult items were removed. Therefore, the actual maximum possible score is 144 instead of 160 points. Moreover, WAT has a high Cronbach's alpha reliability of 0.96, which confirms previous studies' reliability estimate (e.g. Read, 1993, 1995).

Reading Comprehension (RC) is a standardized multiple choice question test used to examine basic comprehension. It is drawn from 'Scholastic Assessment Test' (SAT) in 'Critical Reading for SAT' testing manual (Paul & Roger, 2008, pp. 15-23). RC section in SAT examines academic English for students willing to pursue their studies in US College. The main reason for using this test as an RC estimate is the similarity in terms of length and topic of the two passages it includes. Both are around 552 tokens in length and related to politics. Another reason is that its comprehension questions do not relate to one another, revealing students' actual understanding of

the passages. To investigate the effect of a small percent of unfamiliar words on comprehension, only the first passage (hereafter RC_1) was adjusted. Hence, 18% of its high information value words were replaced with their low frequent synonyms. The objective was to examine the extent to which a low percent of this type of words may influence comprehension and thus estimate text complexity in terms of lexical difficulty, and foreground the importance of teaching this lexical category.

The two passages include multiple-choice questions that are straightforward and literal with a maximum possible score of 17 (nine questions for RC_1 and eight for RC_2). Indeed, there was a need to reduce RC length due to the large number of items in the lexical tests in order to lessen examinees' fatigue and maximize their motivation to answer the three tests items completely. Finally, like the vocabulary tests, RC measure has a high parallel-form reliability of 0.81, indicating estimation of the same skills: text word meaning recognition and comprehension.

Recall is the second sub-section of RC test wherein students are required to write the ideas they retain from each passage to assess their comprehension. Texts analysis into different levels of idea units was based on Zerhouni's hierarchy (1998). Thus, at the highest level are main ideas of the text (allotted 2 points), followed by secondary ideas directly supporting them (allotted 1.5); then tertiary ideas of less importance (allotted 1); and at the last are quaternary ideas which constitute minor ideas (i.e. details) (allotted 0.5) (Zerhouni, 1998, p. 66). This classification helps to determine which type of ideas students with different lexical knowledge understand and recall. In statistical analysis of results, recall of RC_1 and RC_2 were compared to one another and correlated with vocabulary size and depth.

Frequency was estimated to determine the lexical difficulty of both reading passages. Following Chen et al (2016) method, the present study measured the mean and standard deviation of the word-frequency levels in each passage. Prior to this are two steps to report here. First, words were tokenized wherein all function items and proper nouns were eliminated because they are highly frequent and carry less meaning. Passages were then manually entered into software called range (available in Nation's official website) to supply the number of word tokens and types. This operation revealed that RC_1 counted 130-word types and RC_2 145. Word type was used as a basis for this analysis because it provides a better estimate of text difficulty than that of word token (Chen & Meurers, 2017). The second step consisted in matching word types of the passages with the stratified frequency lists of headwords developed from British Nation Corpus (BNC) and Corpus of Contemporary American English (COCA) (Nation, 2012). Indeed, using two corpora that cover a wide range of written as well as spoken lexis in natural language learning enables a better estimate of word frequency (Chen et al., 2016) and receptive lexical knowledge of learners.

Data on frequency estimate is ordinal (i.e. from the highest to the lowest levels). Sums of lexical items in each frequency level were counted and divided by the total number of word types.

For instance, in RC_1 the percentage of words in the 2nd 1000 frequency level is $^{15}/_{130}$= 11.54% whereas in RC_2 it is $^{34}/_{145}$= 23.45%. Figure 1 presents graphically the percentages of

word types across the frequency spectrum. It shows that the second passage contains a large number of high-frequency words and low percentages of low frequency words starting from the fourth frequency level, RC_2 (*M*= 1.64, *SD*= 1.12). The first passage, on the other hand, consists of more low frequency words, RC_1 (*M*= 1.98, *SD*= 1.84). The means and standard deviations presented here reveal that RC_1 is lexically more difficult than RC_2.

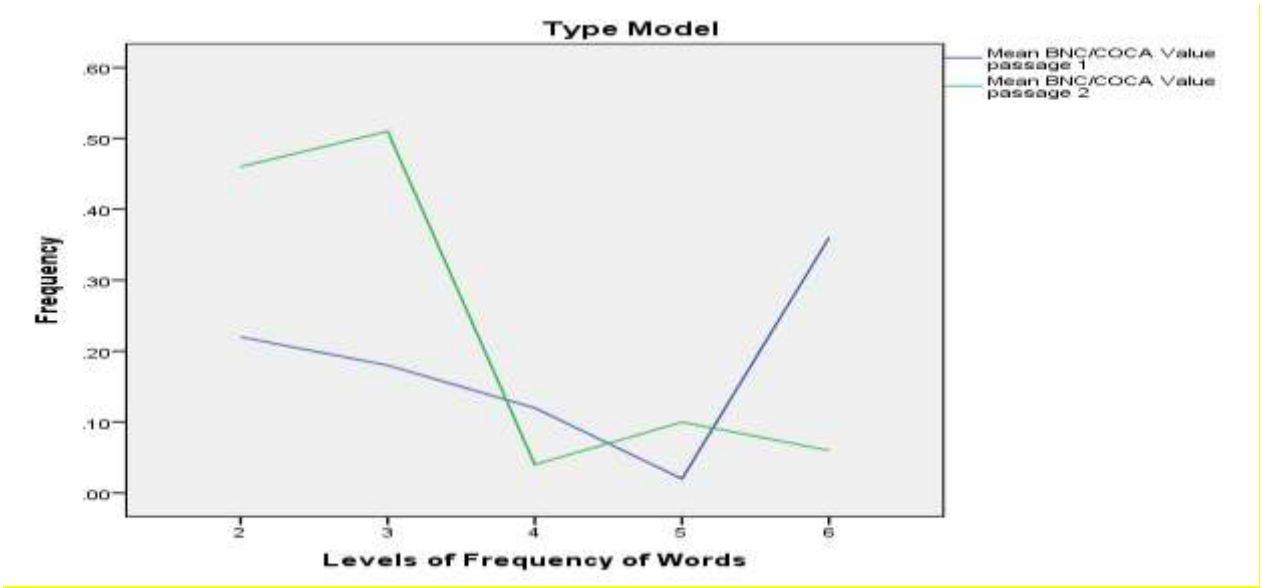

Figure 1. Word frequency percentages of passage 1& 2

Procedure

Test administration took place before the final exams of the academic year (March 2016). Prior to testing, professors were briefed about the objectives of the study and the instruments used to collect data. Students were informed then about the reasons of the researcher's visit and the importance of participating in this study to contribute to research. Students willingly accepted to take the tests for their curiosity to find out the quantity and quality of their vocabulary knowledge. Besides, the testing maximum time was two hours, and the number of students in every session - ranging from five to 17- was manageable for supervision with a total of six sessions. All tests were in paper-and-pencil format and students were allowed to take five minutes break in between tests.

Data Analyses

Data collected from the three instruments, namely VST, WAT, and RC is interval in nature. This is because the present study attempts to look into the effect and type of relationships among variables. Data was analyzed through SPSS 22 software. Both descriptive and inferential statistics were run. Inferential statistics include paired samples t-test to check the significance level of the effect of lexical difficulty on comprehension and recall, and Pearson product-moment correlation to estimate the relationship between scores on vocabulary size, depth, comprehension, and recall. For the purposes of the present study, scores of the three tests were converted into 100 to be evenly analyzed.

Results

Table 1 shows the descriptive statistics of the variables. The first row concerns the vocabulary size test (VST); the second row is devoted to vocabulary depth (WAT); the third and fourth rows (RC_1 and RC_2) refer to the first and the second sub-measures of reading comprehension respectively; and the fifth and sixth rows are devoted to recall of passage 1 and 2.

Table 1. *Means, standard deviations, minimum and maximum of all research variables (N=80)*

	Minimum	Maximum	Mean	SD
VST	30.00	93.00	62.88	16.01
WAT	29.86	88.89	61.97	14.50
RC_1	11.11	67.67	39.32	15.85
RC_2	12.50	100.00	59.53	19.30
Recall 1	10.00	80.00	40.50	17.56
Recall 2	.00	90.00	45.75	20.17

As a reminder, the first objective of the study is to investigate whether lexical difficulty, as estimated by frequency, has an effect on reading comprehension and recall. Frequency variable is analyzed here from the semantic perspective since it has a direct impact on reading comprehension (Marks et al., 1974). As shown in descriptive statistics (Table 1), the mean of RC_1 is lower than that of RC_2, indicating a low performance in the first passage and higher scores in the second one. Similarly, the mean of the first passage recall is lower than that of the second passage. However, it is unclear at this stage whether the difference is statistically significant. Table 2 shows paired samples t-test results where scores on RC_1 and RC_2 as well as the first and the second recalls are contrasted.

The second objective is to estimate the relationship between scores on vocabulary size, depth, comprehension, and recall. Pearson product-moment correlation was used to measure these relationships, with vocabulary knowledge (i.e. size and depth) being the independent variable and comprehension and recall the dependent variables. As shown by research (e.g. Qian, 1998, 2002), vocabulary knowledge has a close relationship with reading, however, it is unclear which aspect of lexical knowledge is more required to ensure comprehension. Analyzing the variables' connections helps also to determine the lexical threshold Moroccan EFL BA (Bachelor of Arts) learners reach. In fact, this paper focuses on basic comprehension to foreground previous findings relative to this question. Table 3 presents the results concerning this objective.

Table 2. *Paired samples t-test of RC and recall (N=80)*

	M	t	df	Sig (1-tailed)
RC_1- RC_2	-20.20160	-12.892	79	.001
Recall 1- Recall 2	-5.25000	-2.651	79	.01

Significance level $p< 0.01$ (one-tailed)

Regarding the first objective of the study, a one-tailed paired samples t-test revealed that students scored less in the first sub-test of reading comprehension (RC_1) (*M* = 39.32, *SD*= 15.85) compared to their performance in RC_2 (*M*= 59.53, *SD*= 19.30), t(79)= -12.892, *p*<.001. Results, in Table 2, showed a statistically significant difference between participants' performance in RC_1 and RC_2. As mentioned earlier, RC test has a high reliability of 0.81 using parallel form analysis to account for the similarity of the skills estimated in both sub-tests. The two passages differ only in terms of the difficulty created in RC_1 through the substitution of a small percent of frequent lexis of high information level by less frequent synonyms compared to the high frequencies of key words in RC_2. Measurement of the overall lexical difficulty of the passages shows that frequency mean and standard deviation of (RC_1) is (*M*= 1.98, *SD*= 1.84), whereas that of RC_2 is (*M*= 1.64, *SD*= 1.12). This indicates that RC_1 is more demanding than RC_2 given the lexical difficulty of the first passage. Stratifying frequency levels of the words into bands (i.e. high, mid, and low) gives an insightful description (Schmitt, 2014). It confirms that the first passage has more mid-low frequency words than the second passage.

As also revealed by t-test results of free recall task (Table 2), participants recalled fewer ideas from RC_1 (*M*= 40.50, *SD*=17.56) compared to RC_2 (*M*= 45.75, *SD*=20.17), t(79)= -2.651, *p*<.01. RC_1 recall protocols included minor details but fewer major points due to the difficult words substituted for the passage high information level words which, most probably, hindered comprehension.

Table 3. *Correlations between VST, WAT, $RC_{1\&2}$, and recall 1&2 (N=80)*

Test	VST	WAT
VST		
WAT	.677**	
RC_1	.717**	.440**
RC_2	.594**	.384**
Recall 1	.499**	.506**
Recall 2	.590**	.568**

** Significance level $p< 0.01$ (one-tailed)

Concerning the second objective, Table 3 shows high, moderate to low correlations among variables. To begin with the results of lexical knowledge, tested through vocabulary size (VST) (*M*=62.88, *SD*=16.01) and vocabulary depth (WAT) (*M*=61.97, *SD*=14.50), Pearson's *r* data analysis revealed a significantly moderate positive correlation, *r*= .677, *p*< .01. That is, students who scored high in vocabulary size had somewhat high scores in depth and vice versa. As for the relationship between VST and RC_1 (*M*=39.32, *SD*= 15.85), the correlation coefficient was significantly high, *r*= .717, *p*< .01, whereas it was remarkably low, *r*= .440, *p*<.01 for the WAT - RC_1 relationship. Lower correlation coefficients were also found between the two lexical aspects and the second reading sub-measure. VST and RC_2 (*M*= 59.53, *SD*= 19.30) had a moderate positive correlation, *r*= .594, whereas relationship between WAT and RC_2 revealed a significantly low correlation, *r*= .384. These results show that VST correlates higher with RC_1 and RC_2 whereas WAT lags behind considerably, and that performance in RC_1 required more lexical competence compared to RC_2.

As exhibited by table 3, recall had significantly moderate to low correlations with VST and WAT. A Pearson product-moment *r* analysis revealed a low positive correlation between VST and recall 1 (*M*= 40.50, *SD*=17.56), *r*= .499, *p*< .01. Interestingly, the correlation between WAT and RC_1 recall was higher, with a coefficient of *r*= .506. On the other hand, scores on RC_2 recall (*M*= 45.75, *SD*= 20.17) correlated moderately with VST and WAT, *r*= .590, *r*= .568, *p*<.01 respectively. To sum up, performance in recall can be ascribed to the degree to which the lexicon used obscures the passage meanings.

Discussion

The present study explored the effect of lexical difficulty estimated by frequency (Chen & Meurers, 2016) on reading comprehension and recall. It also investigated the relationship between vocabulary size, depth, reading comprehension, and recall.

For the first objective, the level of lexical complexity seems to have an effect on reading comprehension and recall. Thus, lexical difficulty hinders both comprehension and recall. Table 2 shows a statistically significant difference between scores obtained in students' performance in passage 1 as compared to passage 2. These results corroborate those of previous studies, which came to the same conclusion (Marks et al., 1974; Anderson et al., 1981). Consequently, the lexical frequency model is confirmed to estimate efficiently text difficulty (Chen et al., 2016, 2017). Unlike Anderson et al. (1981) findings, however, free recall seems to decrease in the lexically difficult passage. This discrepancy is probably due to the variety of information level of substituted words in their experiments.

Thus, the lexicality of a text is fundamental to comprehension, particularly if the proportion of unknown words is of high information value (Stratton & Nacke, 1974; Mezynski, 1983). For this reason, comprehension performance declined in the first passage while it was higher in the second one as table 1 shows. More importantly, skilled readers are known for having high representations of lexis in their mental lexicon (Klare, 1968). These representations are probably the different aspects of words that can be epitomized in high and low frequently used synonyms, as it can be exemplified in word use and its different meanings. Hence, students with poor or low

lexical representations do not exhibit a good control of comprehending texts while reading. Vocabulary learning, therefore, seems to be learners' burden (Nation, personal communication) since it highly influences other language skills such as reading. However, if learning is generally fine-tuned to students' levels and needs, they can achieve better results. For example, estimating the difficulty of academic materials might be the initial solution to resort to at the EFL undergraduate levels. This is not to be considered as an extra load of work for teachers, but a necessary step for the reform of university educational system. Hence, assembling a corpus of estimated reading material can be an effective method to start with.

Recall of the two passages has further confirmed participants' level of understanding and retrieving ideas from the reading material. Thus, students scoring high in vocabulary size and depth recalled more ideas than those with lower scores in the lexical tests. As a reminder, ideas in the recall task have been categorized according to Zerhouni's hierarchy (1998, p.66) (see methods). The order of these ideas is crucial as it demonstrates whether readers have grasped the passage fully or only partially. Therefore, recall is a crucial technique for evaluating comprehension and retrieval of text content. It seems that the more ideas students recall from a passage, the more cognitive effort they make in deciphering word meaning (Anderson & Freebody, 1981).

As for the second objective of the study, it has been shown that vocabulary knowledge has a close relationship with comprehension and recall. This confirms results from previous studies (Qian, 1998, 2002; Zhang, 2012; Horiba, 2012). Interestingly, vocabulary size correlates higher with reading comprehension in both passages compared to vocabulary depth. The latter, however, correlates highly with the recall task. This may indicate that for basic reading comprehension, vocabulary size is more of a powerful lexical component. Qian's (1998, 2002) results may refute this claim since depth was found to be a stronger indicator of reading in his studies. However, one should note that Qian's findings might be due to the effect of depth adjustments. To explain, this researcher substituted 10 words from vocabulary depth test (Read, 1993) with words and their stimuli from the RC test he employed in his investigation. In this case, one cannot disregard the interwoven effect of the two tests ten identical items on students' answer. If lexical items happen to appear twice in different measures, students are likely to learn them and be able to answer correctly. Above all, this may confirm the idea of "pressure point", mentioned earlier, in reading comprehension (Perfetti, 1985; Perfetti et al., 2002; Perfetti, 2007). Though different reading tasks (i.e. comprehension or recall) may necessitate different lexical knowledge, the latter remains a pressure point between word recognition and text comprehension. It can also be interpreted as the transactional phase between the meaning of the text and the meaning understood by the reader through associating text words with his/her mental lexicon (Marks et al, 1974; Ryder & Slater, 1988).

Indeed, qualitative assessment of texts is subjective as it may differ from one reader to another. It is based on an impressionistic view mainly. The conceptual shift from word to text meaning in comprehension requires a quantitative evaluation of text lexis. This is because words can be tokenized and analyzed objectively. As research studies reported above have shown, lexicon is a crucial component for understanding and conveying verbal ideas. Though such a numerical analysis seems to be valid in basic comprehension, it may not be sufficient in other comprehension

levels (critical reading especially). The reason is that further factors such as the writer's intentions and the socio-cultural context of text events have to be taken into consideration in assessing reading materials level of difficulty in congruence with learners' reading proficiency level. Thus, demanding comprehension levels necessitate complex analyses in order to come up with different assessment techniques for a better evaluation of students' reading comprehension deficiencies.

Implications and Conclusion

As findings of the present study show, text complexity can be estimated from a lexical perspective using a frequency model. This indicates the crucial role of vocabulary, especially semantic frequency, in reading comprehension and recall. It is therefore essential for EFL teachers to allocate some time to systematically teach and test vocabulary, and to teach words in association since they are stored in the mental lexicon as a network. Besides, given the semantic load of lexical items, students should be exposed both to their high and low synonyms. More importantly, estimating text difficulty for different academic levels is crucial for selecting reading materials whose degree of lexical difficulty is in accordance with students' levels. Similarly, it is also recommended that students applying for English departments sit for a standardized lexical test as an effective assessment tool for their placement. Another suggestion based on this study findings is to provide EFL University students with training in reading and vocabulary learning strategies in order to equip them with helpful tools to overcome anxiety toward reading in general, better benefit from their reading lessons and assignments, and to systematically expose them to a wide range of lexis. Additionally, an EFL corpus of estimated reading material should be assembled and put at the disposal of teachers. Finally, the present investigation was limited to basic comprehension; thus, measuring the frequency effect on higher levels of comprehension is recommended to fully grasp this issue.

About the Authors:

Nadia Nouri is a Ph.D. candidate in "Applied Linguistics and TEFL" research unit at the Doctoral Center of the Faculty of Education, Mohammed V University-Rabat, Morocco. She holds a master degree in Teaching English as a Foreign Language (TEFL) from Ibn Tofail University, Kenitra-Morocco, and is currently teaching English at the American Language Center in Rabat and at the INPT (Posts and Telecommunication National Institute).
ORCid ID: 0000-0001-9775-3102

Pr. Badia Zerhouni (Ph.D. holder from Ottawa University) is an associate professor of Applied Linguistics at the Faculty of Education, Rabat, Morocco, and former director of the Doctoral Center of the Faculty of Education. She currently teaches and supervises research projects in MA Applied Linguistics /ELT programs and supervises doctoral research in different subjects related to AL/ELT. Her research interests include reading, writing, and vocabulary acquisition in EFL contexts.

Acknowledgment

We would like to thank Dr. Paul Nation for his help in the present project, which is part of a doctoral dissertation.

References

Adams, M., & Bruce, B. C. (1980). *Background Knowledge and Reading Comprehension*. Retrieved from https://www.researchgate.net/publication/234650027_Background_Knowledge_and_Reading_Comprehension_Reading_Education_Report_No_13

Alderson, J. C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.

Anderson, R. C., & Freebody, P. (1979). *Vocabulary Knowledge and Reading*. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/17498/ctrstreadeducrepv01979i00011_opt.pdf?sequence=1

Anderson, R. C., & Freebody, P. (1981). Vocabulary Knowledge. In J. T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews* (pp. 77-117). Newark, DE: International Reading Association.

Bernhardt, E. B. (2005). Progress and Procrastination in Second Language Reading, *Annual Review of Applied Linguistics*, *25*, 133-150.

Chen, X., & Meurers, D. (2016). Characterizing Text Difficulty with Word Frequencies. Proceedings of *the Eleventh Workshop on Innovative Use of NLP for Building Educational Applications*. San Diego: California.

Chen, X., & Meurers, D. (2017). Word Frequency and Readability: Predicting the Text-Level Readability with a Lexical-Level Attribute, *Journal of Research in Reading*, 1–25.

Ellis, N. C. (2012). Frequency-based Accounts of Second Language Acquisition. In S. M. Gass & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 193-210). USA: Routledge.

Henriksen, B. (1999). Three Dimensions of Vocabulary Development, *Studies in Second Language Acquisition*, *21*, 303-317.

Horiba, Y. (2012). Word Knowledge and its Relation to Text Comprehension: A Comparative Study of Chinese- and Korean-Speaking L2 Learners and L1 Speakers of Japanese, *Modern Language Journal*, *96*, 108–121.

Hu, M., & Nation, I. S. P. (2000). Unknown Vocabulary Density and Reading Comprehension, *Reading in a Foreign Language*, *13*, (1), 403−430.

Just, M. A., & Carpenter, P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension, *Psychological Review*, *87*, (4), 329–354.

Klare, G. R. (1968). The Role of Word Frequency in Readability, *Elementary English, 45*, (1), 12-22.

Koda, K. (2005). Insights into Second Language Reading: A Cross-Linguistic Approach, *Journal for Language Learning*, *24*, (2), 88-92.

Laufer, B. (1989). What Percentage of Text-Lexis is Essential for Comprehension? In L. Ch & M. Nordman. (Eds.), *Special Language: From Humans Thinking to Thinking Machines* (pp. 316-323). Multilingual Matters.

Laufer, B. (1992). How much is Lexis Necessary for Reading Comprehension? In H. Bejoint & P. Arnaud (Eds.) *Vocabulary and Applied Linguistics* (pp. 126–132). Macmillan, Basingstoke & London.

Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness, *Language Learning*, *54*, 399-436.

Lexile. (2007). *The Lexile Framework for Reading: Theoretical Framework and Development*. Technical Report, MetaMetrics, Inc., Durham, NC.

Lively, B. A., & Pressey, S. L. (1923). A Method for Measuring the Vocabulary Burden of Textbooks. In W. H. DuBay, (Ed), *the Classic Readability Studies*, (pp. 6-16). BookSurge Publishing.

Marks, C. B., Doctorow, M. J., & Wittrock, N. C. (1974). Word Frequency and Reading Comprehension, *Journal of Educational Research*, *67*, 259-262.

Martin, C. J., & Pantalion, C. A. J. (1973). Effect of two Word Deletion Schemes upon the Comprehension of Five Levels of Telegraphic Prose, *The Annual Meeting of the American Educational Research Association*. New Orleans.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix.* New York, NY: Cambridge University Press.

Mezynski, K. (1983). Issues Concerning the Acquisition of Knowledge: Effects of Vocabulary Training on Reading Comprehension, *Review of Educational Research*, *53*, 253-279.

Milone, M., & Biemiller, A. (2014). *The Development of ATOS: The Renaissance Readability Formula*. Technical Report, Renaissance Learning, Wisconsin Rapids.

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters.

Nacke, P. L. (1970). *Skimming Strategy in Reading as a Function of Familiarity with Content and Redundancy Reduction in Printed Discourse*. (doctoral dissertation). Retrieved from https://open.library.ubc.ca

Nation, I. S. P. (2001). How Many High-Frequency Words are there in English? In M. Gill, A. W. Johnson, L. M. Koski, R. D. Sell & B. Warvik (Eds.), *Language, Learning, and Literature:* Studies Presented to Hakan Ringbom English Department Publications (pp.167-181) Abo Akademi University, Abo.

Nation, I. S. P. (2006). How Large is a Vocabulary Needed for Reading and Listening? *Canadian Modern Language Review*, *63*, (1), 59-82.

Nation I. S. P., & Beglar, D. (2007). A Vocabulary Size Test, *the Language Teacher*, *31*, (7), 9-13.

Nation, I. S. P. (2012). The Vocabulary Size Test. Retrieved from https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of Text Difficulty: Testing their Predictive Value for Grade Levels and Student Performance*. Technical Report, the Gates Foundation.

Ojemann, R. J. (1934). The Reading Ability of Parents and Factors Associated with Reading Difficulty of Parent Education Materials. In W. H. DuBay, (Ed), *the Classic Readability Studies* (pp. 27-30). BookSurge Publishing.

Paul, L., & Roger S. W. (Eds). (2008). *Critical Reading for SAT*. USA: Peterson.

Patty, W. W., & Painter, W. I. (1931). A Technique for Measuring the Vocabulary Burden of Textbooks. In W. H. DuBay, (Ed), *the Classic Readability Studies* (pp. 31-40). BookSurge Publishing.

Perfetti, C. A. (1985). *Reading Ability*. New York, NY: Oxford University Press.

Perfetti, C. A., & Hart, L. (2002). The Lexical Quality Hypothesis. In L. Verhoeven. C. Elbro, & P. Reitsma (Eds.), *Precursors of Functional Literacy* (pp. 189-213). Amsterdam: John Benjamins.

Perfetti, C. A. (2007). Reading Ability: Lexical Quality Hypothesis, *Scientific Studies of Reading, 11*, (4), 357-383.

Qian, D. (1998). *Depth of Vocabulary Knowledge: Assessing its Role in Adults' Reading Comprehension in English as a Second Language* (Unpublished doctoral dissertation). University of Toronto, Canada.

Qian, D. (2002). Investigating the Relationship between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective, *Language Learning, 52,* 513-536.

Rayner, K., & Duffy, S. A. (1986). Lexical Complexity and Fixation Times in Reading: Effects of Word Frequency, Verb Complexity, and Lexical Ambiguity, *Memory & Cognition*, *14*, (3), 191–201.

Read, J. (1993). The Development of a New Measure of L2 Vocabulary Knowledge, *Language Testing, 10*, 355-371.

Read, J. (1995). Validating the Word Associates Format as a Measure of Depth of Vocabulary Knowledge. *Paper Presented at the 17th Language Testing Research Colloquium*, Long Beach, CA.

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the Depths: How should the Construct of Vocabulary Knowledge be Defined? In B. Harley, & J. H. Hulstijn (Eds.), *Vocabulary in a Second Language:* Selection, Acquisition, and Testing (Vol. 10, pp. 209-227). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ryder, R. J., & Slater, W. H. (1988). The Relationship between Word Frequency and Word Knowledge, *Journal of Educational Research*, *81*, (5), 312-317.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows, *Language Learning*, *64*, (4), 913–951.

Stratton, R. P., & Nacke, P. L. (1974). The Role of Vocabulary Knowledge in Comprehension. In P. L. Nacke (Ed), *23rd Yearbook of the National Reading Conference*, Clemson, S. C: National Reading Conference.

Stuart, M., & Kramer, B. (2015). The Creation of a New Vocabulary Levels Test, *Shiken*, *19*, (2), 1-11.

Zerhouni, B. (1998). *Reading in L1, L2, and FL: A Variable-Treatment Study*. (doctoral dissertation). National Library of Canada, Ottawa.

Zhang, D. (2012). Vocabulary and Grammatical Knowledge in Second Language Reading Comprehension: A Structural Equation Modeling Study, *Modern Language Journal*, *96*, 558–575.