

Amount of Vocabulary Required to Grasp Comprehension in a Text

Jalal H. Albaqshi

General Studies, Alahsa College of Technology,
Alahsa, Saudi Arabia

ABSTRACT

This study addresses the relationship between vocabulary and reading comprehension and Showing some other factors which affect the comprehension beside vocabulary. A descriptive analysis and correlation test were applied to show why some L2 learners cannot understand a gist of a text though they are familiar to many vocabulary. The sample group in this study consisted of 64 male Saudi learners. They were between the ages of 18 to 22 years old and had been studying English in formal settings for six years in the form of moderate doses of input but not intensive English. This study highlights the percentage coverage of vocabulary in a text rather than the whole learners 'reservoir of vocabulary in order to understand the general ideas of a text. Results show that L2 learners need to be familiar at least with 62% to 71% of vocabulary in a text to grasp the gist of whole passage. This result helps to write down several implications for researchers and teachers of English as a second and foreign language.

Keywords: streaming vocabulary, keyword vocabulary, reading comprehension.

INTRODUCTION

Reading comprehension is mutually interwoven with vocabulary knowledge. Studying the relationship between understanding a text and the amount of known vocabulary is fundamental to realizing how one affects the other, if at all. In fact, vocabulary is a major constituent of reading comprehension, while reading is considered a primary source of vocabulary acquisition. This relationship leads to a common, logical belief which implies the more vocabulary a learner knows, the better he/she is able to understand a text. However, the process of linguistic development is more complicated than word recognition and understanding. Several factors can also play roles in communicating with a written text. This fascinating relationship adds viability for both linguistic components to contribute to linguistic competence. Also, other influential aspects in the relationship between comprehension skills and vocabulary knowledge may be considered to overcome some of the challenges of reading comprehension. When Scrivener (2009) discussed reasons for difficulties with reading in a foreign language, he mentioned lack of vocabulary, slow reading, and inability to understand all vocabulary. Essentially, according to Scrivener, the vocabulary component is crucial for the entire reading process. It is important to consider the possibility that the words of a text may be familiar for a learner, but he/she may still be unable to understand the whole message. This is related to structural and pragmatic aspects which can be mastered gradually in higher level of acquisition. Some other factors can be added as influential:

critical thinking development, cognitive aspect, and schemata. All mentioned factors play considerable roles in the effectiveness of processing texts to produce logical conclusions which match the author's intended message.

A focal issue that should be considered when examining the relationship between reading and vocabulary is the following question: *Should researchers highlight the amount of vocabulary recognized (known) in a text by an L2 learner or the total amount of vocabulary an L2 learner knows?* Richards (2002) discussed dilemmas for second language reading instruction. One dilemma directly related to our highlighted issues in this study is the many different contexts and genre structures that L2 learners encounter. This dilemma drives us to emphasize vocabulary recognition percentage in a text rather than the size of a learner's whole reservoir of vocabulary. Let us suppose a learner who has sufficient vocabulary to be fluent in English reading is exposed to a text such as an article in an economic magazine. Would this learner understand the text as much as the other general English texts? S/he would most likely not, because there would be more complicated specialized terms and vocabulary which belong to a certain genre.

LITERATURE REVIEW

The relationship between vocabulary size and reading comprehension is a common topic which has been attracting researchers for more than two decades. For example, the focus of a study conducted in Malaysia by Kameli and Baki (2013) was the relationship between vocabulary breadth/size and reading comprehension. The methodology used involved administering a vocabulary test designed by Nation and an IELTS reading test. The data were analyzed in terms of correlation coefficient (r), mean, and "p" value, and the results showed a high level of statistical significance between vocabulary familiarity and reading comprehension. Another study conducted by Farvardin and Koosha (2011) involved the examination of the relationship between depth and breadth of vocabulary and reading comprehension. They submitted reading comprehension test and vocabulary language test 1000, 3000, and 10000. The results showed a positive relationship between the two aspects of breadth and depth and reading comprehension.

Another study was conducted by Gue and Roehrig (2011), who examined the roles of metacognitive awareness of reading strategies, syntactic awareness in English, and English vocabulary knowledge in the English reading comprehension of Chinese-speaking university students ($n = 278$). The results suggested a two-factor model of a General Reading Knowledge factor (metacognitive awareness employed during the English reading process) and a Second Language (L2) Specific Knowledge factor (comprising vocabulary knowledge and syntactic awareness) offered the best fit to the data; 87% of the variance in reading comprehension was explained by the two factors together. L2 Specific Knowledge was a stronger predictor of reading comprehension than metacognitive awareness. A multigroup analysis was conducted using structural equation modeling to compare poor-reader and good-reader groups. The correlation between L2-specific knowledge and metacognitive awareness and their relation to reading comprehension was the same across groups. It should be noted that most studies involved examination of each learner's vocabulary reservoir rather than the recognized words of a text. This draws the attention toward a learner's reading levels rather than his/her focus on a text and its content in relation to his/her cognitive communication via written codes. A group of studies, such as those conducted by Pang (2008), Kameli and Baki (2013), and Laufer (2010), highlighted vocabulary size rather than text coverage

to gauge each reader's understanding of a text they had read. Nation (2006), in one of his articles, focused on vocabulary size but also mentioned text coverage for comprehension.

RESEARCH METHODOLOGY

Descriptive analysis showing the data distribution in quadruple classification was used for this study. After data collection, students' responses were collected and classified as four areas: area a, area b, area c and area d. These areas are displayed in a chart below to show the effect of vocabulary size on reading comprehension. We hypothesized that our results would indicate the existence of a strong relationship between vocabulary and reading comprehension; therefore, correlation is included in results. In addition, displaying a linear relationship between the two variables is addressed.

Research Questions

1. What is the minimum percentage of familiar streaming and keywords needed to comprehend a text?
2. Why do some learners have low comprehension levels though they have high vocabulary recognition?

DATA COLLECTION AND SAMPLING

The sample group in this study consisted of 64 male Saudi learners. They were between the ages of 18 to 22 years old and had been studying English in formal settings for six years in the form of moderate doses of input but not intensive English. The sampling method used for the study was systematic sampling — every individual was a potential participant. Data was collected with the aid of a test which consisted of a short passage, and students were asked to underline each word that was unknown to them. Then they read the whole passage and did their best to understand as much as they could. During the final stage, they were asked to write in their own language what they understood from the text.

RESULTS

Figure 1 below shows the results of the participants' responses. The graph is divided into four areas: area a, area b, area c, and area d. As explained in Table 1, area a includes the students whose understanding was low and recognized only a few words from the text. On the contrary, area b includes respondents with high vocabulary recognition and understanding as well. Areas a and b represent the logical concept that a larger vocabulary leads to greater comprehension. Although areas c and d constitute only 27% of the participants, they are in the circle of problematic and questionable results (high vocabulary recognition and low understanding and vice versa).

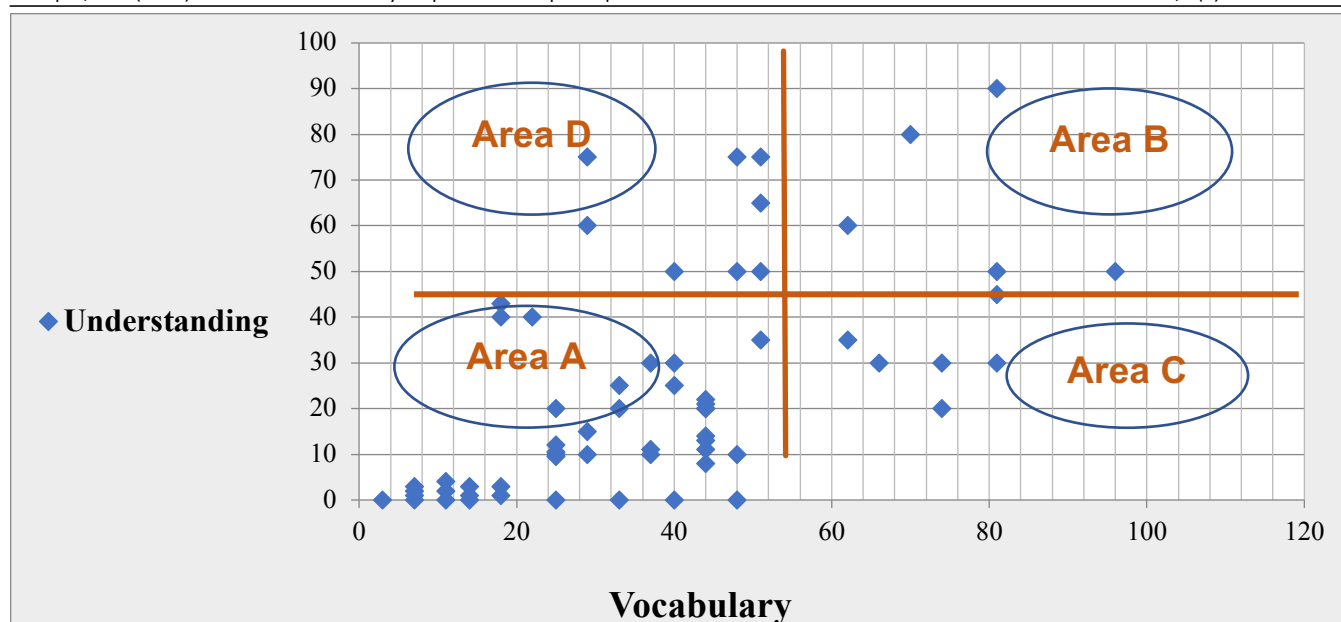


Figure 1

Table 1

Area	Learners' Results	Factors affecting
A	Low vocabulary + Low understanding 60%	Lexical factor Linguistic intelligence
B	High vocabulary + High understanding 11%	Cognitive factor
C	High vocabulary + Low understanding 15.8%	Structural aspect Pragmatic aspect
D	Low vocabulary + High understanding 12.6%	Inference Guessing Reflective learners

Table 1 shows the classification of respondents and their percentages as follows: area a (60%), area b (11%), area c (15.8%), and area d (12.6%). Most of the respondents are condensed in area a, and the lowest percentage of respondents fall into area b. The observations which are considered and analyzed here are noticeable when they are in inverse relationship. This can be illustrated as in in Figure 2 below. In Figure 2, there are three highlighted observations (cases of learners) which are worth of discussion. They are observations a, b, and c. There are other observations with similar situations, but none is as noticeable as these three. The three observations a, b and c are in inverse relationship between vocabulary size and comprehension. In observation a, a learner is familiar with 80% vocabulary of the text but understood only around 32% of the text. This can leads us to think about beyond vocabulary effect on understanding. Observation b is in a similar situation of a. However, observation c has high understanding (78%) but low vocabulary recognition (25%).

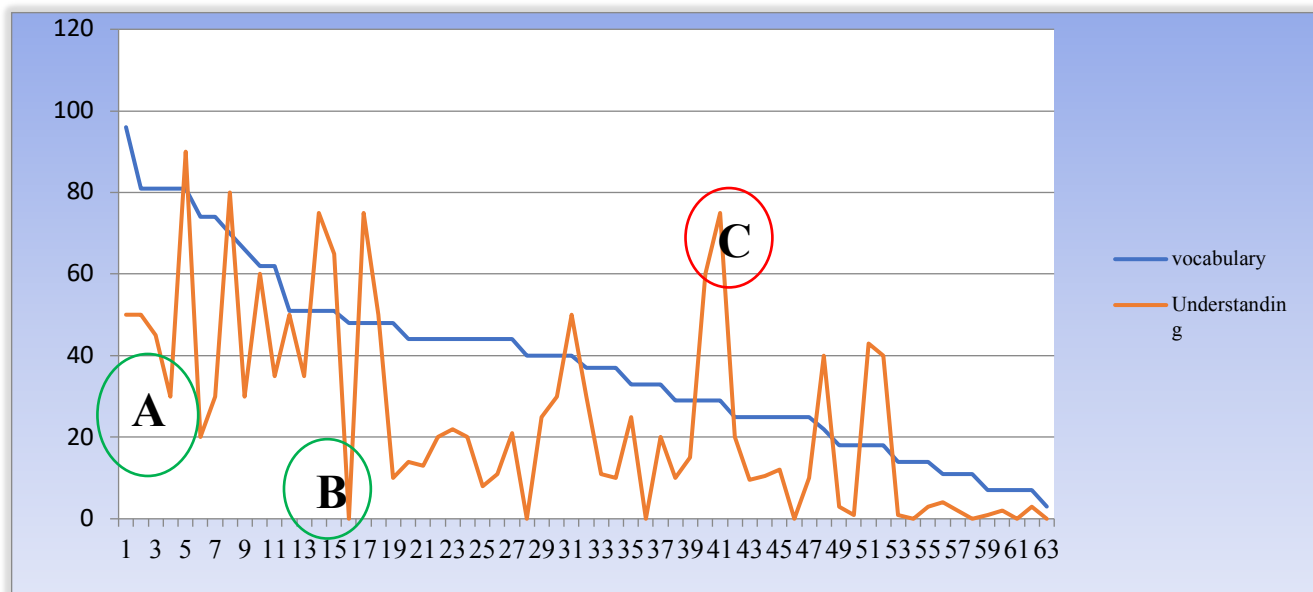


Figure 2: Relationship between vocabulary and understanding

Table 2: Correlation results among comprehension score, stream words and keywords

Correlation			
Variables		Pearson correlation	P-Value
DV	Comprehension score	0.632	0.000
IV	Stream vocabulary recognized		
DV	Comprehension score	0.664	0.000
IV	Keyword recognized		

When Pearson correlation was processed, as the results in Table 2 indicate, a relatively considerable correlation was identified between keyword recognition and understanding with 0.664 and 0.632 for comprehension and streaming words.

DISCUSSION

The results of this study indicate that there are four cases regarding the relationship between the amount of vocabulary a learner recognizes in a text and his/her level of understanding of the text. Some learners show high understanding as a result of high recognition of the vocabulary in the text. Such recognition enables learners to decode the message of a text; however, this group of learners constituted only 11% of the sample (as seen in Table 1). In contrast, 60% of participants could not decode the message effectively due to the lack of vocabulary mastery. Up until this point, we are at the logical baseline because more vocabulary facilitates greater comprehension as words are a basic part of the linguistic components in a text. However, the learners with high vocabulary recognition and low understanding can be put in a speculative and questionable stance. This can be caused by cognitive factors, automaticity and lack of reading strategies employment. Stephen D. Krashen (1987) suggested that, in order to help a learner to progress in language learning, one should provide comprehensible input. This can be achieved in different ways:

1. A: slower rate and clearer articulation
2. B: more use of high frequency vocabulary

3. C: syntactic simplification (shorter sentences)

Upon examination of the text which acted as a test in this study, it can be said that it conforms to Krashen's suggestions in which it was conducted in written form and learners were given ample time and never asked to hand in their papers. Also, the text was carefully selected to contain general and high frequency vocabulary (colors). After the test was over, the researcher made sure that the idea of the topic was familiar to the learners' background and that they already knew in their native language the information mentioned in the text. Therefore, students of areas a and b (as shown in Table 2) met Krashen's conditions.

Table 3

Area	Learners' Results
A	Low vocabulary + Low understanding 60%
B	High vocabulary + High understanding 11%
C	High vocabulary + Low understanding 15.8%
D	Low vocabulary + High understanding 12.6%

Based on input hypothesis, Krashen (2003) asserted that L2 learners need comprehensible input to progress with second language acquisition, and the input should not be only comprehensible but also sufficient, as in the formula "*i + 1 is present*". This hypothesis motivates and gives this study its significance in which researchers and English teachers need to realize how much familiar vocabulary should be in a text in order to help a learner develop his/her skills in reading. Full familiarity of words in a target text would contradict the *i + 1 hypothesis*, while a text including too many unknown words in a target text would lead to deficiency in comprehension, frustration and eventually failure. That gives this study its value and significance in one big question: How much vocabulary recognition is needed in order to comprehend a text? Based on the results of this study, a learner needs to be familiar with at least between 62% and 71% of the vocabulary in a text to obtain a fair comprehension of its main ideas and general concepts along with a few details. This result leads to more speculation related to Krashen's input hypothesis. Can we suppose that *i + 1* means that a text is introduced to a learner with familiarity of vocabulary and grammatical rules between 55% to 60 %? Yes, it is true that comprehensible input would facilitate learning, but other factors must be present in order to put together the pieces of the puzzle. One of these factors is learning style, such as field-dependent and field-independent processing. Synthesizing parts and pieces together to formulate a whole spectrum of a picture and understanding relationships of scattered pieces in randomly formed pattern can be a trait of a field-independent (FI) learner whom Brown (2002) describes in his explanation, while field-dependent (FD) describes a learner who can see the whole picture and barely analyzes smaller pieces from the entire frame. FI learners are more likely to use guessing and inferences in their reading of bottom-up processing where FD learners may struggle without enough vocabulary to decode messages.

Therefore, participants who had a higher understanding of the text with less vocabulary recognition might be FI learners. This could be processed inductively — starts from specific to general, from word to phrase, and ending up with sentences and texts. This can be obvious from the sequence of language learning development suggested by Pienemann and Håkansson. Pienemann and Håkansson (1999) explained the sequence in which process skills develop in language learning. This sequence includes five stages:

1. **Lemma:** Words are processed but do not yet carry any grammatical information. Our participants from areas a and c can be categorized in this stage and may barely pass it.
2. **Category procedures:** Lexical items are categorized, and grammatical representation may be added. Our participants from area b can be categorized in this stage.
3. **Phrasal procedures:** Operations at phrase level occur.
4. **S-procedure:** Grammatical information may be exchanged across phrase boundaries. Our participants from area d can be categorized in this stage due to their advanced linguistic development.
5. **Clause boundary:** Main and subordinate clause structures may be handled differently (as cited in Troike 2006 , p. 77).

Background knowledge and understanding the world (schemata) can be another factor which enhances greater comprehension. Horwitz (2008) elaborates in his discussion that reading can be classified into two strategies: phonics-based or whole language approaches. A phonics-based approach depends on bottom-up processing while whole approaches emphasize top-down processing by using background knowledge to understand the smaller pieces of a text and details. It can be logically stated that, with a foreign language, both strategies are useful according to the level of the learners. Beginners utilize a phonics-based approach more than any other strategies. But schemata (whole language approach) is often more compatible for adults since they have more life experience and knowledge, as with our study participants. Brown (1994) clarified the distinction between bottom-up and top-down processing of data: Bottom-up processing that starts from smaller linguistic units toward larger sentences and paragraphs needs sophisticated knowledge of the language itself to perceive the data given. This can answer the question of why some learners have low comprehension though they have good vocabulary recognition . They need to be competent in perceiving data and form a global concept based on the small segments of a text. As Christine Nuttall (1996, pp. 16-17) described, the distinction between them is as follows: bottom-up is comparable to a microscope view while top-down can be like an eagle's view of a landscape. Since our participants were in the range of beginners, they used the bottom-up method when they read so they could try to understand the text from small words and pieces of linguistic representations. For some of them, this may have impeded the flow of the holistic picture of the message in a text.

In Pienemann and Håkansson's sequence, words are an initial stage which, by itself, cannot be considered as a generator of comprehensive understanding of a message within a text. This sequence explains our results, especially in areas a and b in which more vocabulary acquisition led to greater understanding. However, the process should move into the second stage in which grammatical aspect is involved to reach stage three for a phrasal representation. After that, sentence level comes up to be based on words connected with the aid of certain semantic and pragmatic awareness. This linkage was assured by Jeffrey Elman and his colleagues (as cited in Lightbown & Spada, 2006, p. 23). They explained language acquisition in terms of links and connections between words and phrases and where they occur. These links and connections are significant initial steps to comprehension. Cognitive ability is required to triangulate the process from one word to another word to an actual life situation. These abilities of cognition and linking linguistic segments to formulate a meaning was an obstacle for those with low understanding and high vocabulary recognition. Lightbown and Spada (2006) elucidated the connectionist model in which language acquisition is not just a process of associating words and phrases with reality but also a process of

how words and phrases are associated and related to each other (p. 24). We can add to this triangulation the syntactic and pragmatic aspects in which there are more chances for guessing strategies and conclusion. Therefore, a learner needs to employ these aspects while reading a text. Moreover, the level and type of a text play considerable roles in learning development. Indeed, there should be a concordance between a learner's cognitive abilities and the level of the input provided with which to communicate with a text. In addition to age and level of a text, automaticity is significant in learning and development processes. This is related to paying attention and time spent in learning a language. According to the cognitive psychologist, Norman Segalowitz (as cited in Lightbown & Spade, 2003, p. 39), an information processing model is based, in linguistic development, on noticing and "paying attention" to aspects of linguistic representation. However, there is a limit to how much information a learner can pay attention to. This conforms to linking information into a single idea but cannot be achieved in a particular level of development until learners reach a higher level of attention and processing when proceeding to automaticity. Automaticity involves, according to Brown (1994), "a timely movement of the control of a few language forms into the automatic, fluent processing of a relatively unlimited number of language forms" (pp. 64). This explains those participants who achieved low scores in comprehension but recognized a considerable amount of vocabulary (area c). It is worth mentioning one more factor in which, as Brown described, learners are either reflective or impulsive in their learning styles. The former depends on systematic steps in reading and performing mental tasks, while the latter is a risk-taker and depends on guessing. Reflectivity results in accuracy but slower performance. So risk takers may achieve more even if they are not very accurate in their performance.

Reading comprehension requires not only vocabulary recognition but also reading strategies and skills to be implemented while communicating with a text. Lindsay and Knight (2006) listed reading skills learners need to develop. In here, we will examine some focal ones which affected our participants' results. They are as follows:

1. **Understanding the relationship between sentences.** This skill was focal and essential in our study because it acted as a rationale for the area c participants depicted in Table 2 (High vocabulary + Low understanding 15.8%). These students had good vocabulary recognition but low comprehension. Conceiving a relationship between sentences and knowing how a sentence could affect the meaning of subsequent or preceding one is an integral part of grasping the frame of a text's message.
2. **Guessing meaning.** In fact, two areas in Table 2 were affected by this skill. Learners from areas a and d demonstrated low vocabulary recognition but varied in terms of understanding level. Only 12% were able to guess the meaning of words, while 60% were unable to skillfully guess words from context.
3. **Background knowledge of the text.** The researcher made sure that all students were provided with sufficient background on the text. He discussed the topic with participants before the text and discovered that they had mastered the concept of the text.

A question that is always asked is: why are some learners more successful than others? As stated by Troike (2006), part of the answer may be embedded in learners' psychological differences. Other differences include age, sex, aptitude, cognitive style, and motivation. Highlighting age and aptitude would be related to our research. Although the participants in this study are adults aged 18 to 22 with the benefit of analytical abilities, pragmatic skills, and real-world knowledge, our results show

that some of them still performed better than others. Therefore, aptitude and cognitive differences might play a primary role in performance among peers.

CONCLUSION AND IMPLICATIONS

Through this study, we have found answers to our research questions: comprehension is considerably related to the size of vocabulary mastered in earlier stages. Moreover, fewer known words can distort the direction of understanding. Based on our study and results, we can assert that a learner needs to recognize at least between 62% and 71% of the vocabulary in a text to obtain a fair comprehension of its main ideas and general concepts of a text. Thus, we can count on the fact that word recognition can be a compass for understanding a text. However, there are some learners who show different results from the hypothesis. These students might be highly competent with regards to other linguistic aspects, and this may have helped them to comprehend the text's message with the help of only a few words. These aspects could be cognitive aspects, grammatical awareness, and inference and conclusion strategies. In addition, keywords are essential to be recognized.

In terms of pedagogical implications, teachers should be aware of the importance of presenting the factual background of a topic to the learners as well as employing bottom-up strategies (words to phrase to clause to sentence level), and concentrating on word association for effective training on how to link semantic representations together in order to understand the "big picture". However, for field independent learners, it is better to align an optimal strategy to their needs in order to develop their linguistic competence based on their cognitive abilities.

References

- Brown, H. Douglas. *Principles of Language Learning and Teaching*. White Plains, NY, Longman, 2002.
- Brown, H. Douglas. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Englewood Cliffs, N.J.: Prentice Hall Regents, 1994. Print.
- Curzan, Anne, and Michael Adams. *How English works a linguistic introduction*. Harlow: Pearson Education Limited, 2014. Print.
- Elman, J. L., E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Farvardin, Mohammad Taghi, and Mansoor Koosha. "The Role of Vocabulary Knowledge in Iranian EFL Students' Reading Comprehension Performance: Breadth or Depth?" *TPLS Theory and Practice in Language Studies* 1.11 (2011): 1575-580. Web.
- Guo, Ying, and Alysia D. Roehrig. "Roles of general versus second language (L2) knowledge in L2 reading comprehension." *Reading in a Foreign Language* 23.1 (2011): 42-64. Web.
- Horwitz, Elaine Kolker. *Becoming a Language Teacher: A Practical Guide to Second Language Learning and Teaching*. Boston: Pearson/Allyn and Bacon, 2008. Print.
- Kameli, Shima, and Roselan Baki. "The Impact of Vocabulary Knowledge Level on EFL Reading Comprehension." *International Journal of Applied Linguistics & English Literature* 2.1 (2013): 85-89. Web.
- Krashen, Stephen. *Explorations in Language Acquisition and Use*. Portsmouth, Heinemann, 2003.
- Krashen, Stephen. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon, 1982. Print.
- Laufer, Batia. "Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension." *Reading in a Foreign Language*, Vol.22, no.1, Apr. 2010, pp.15-30.

Lightbown, Patsy, and Nina Margaret. Spada. *How Languages Are Learned*. Oxford:

Lindsay, Cora, and Paul Knight. *Learning and Teaching English: A Course for Teachers*. Oxford: Oxford UP, 2006. Print.

Nation, I. "How Large a Vocabulary is Needed for Reading and Listening?" *Canadian Modern Language Review*, vol.63, no.1, 2006, pp. 59-82., doi: 10.3138/cmlr.63.1.59. Oxford UP, 2006. Print.

Pang, Jixian. "Research on Good and Poor Reader Characteristics: Implications for L2 Reading Research in China." *Reading in a Foreign Language*, vol. 20, Apr. 2008, pp. 1-18.

Pienemann, M. & Håkansson, G. (1999). A Unified Approach toward the development of Swedish as L2: A processability account. *Studies in Second Language Acquisition*, 21 (2): 179-420.

Richards, Jack C. & Renandya Willy. *Methodology in Language Teaching: An Anthology of Current Practice*. New York: Cambridge UP, 2002. Print.

Scrivener, J. (2009). *Learning teaching* (2nd ed.). Malaysia: McMillan

Segalowitz, N. 2003. 'Automaticity' in C. J. Doughty and M. H. Long (eds.): *The Handbook of Second Language Acquisition*. Oxford: Blackwell. Pp.382-408.

Troike, Muriel. *Introducing second language acquisition*. Cambridge: Cambridge U Press, 2010. Print.

Ur, Penny. *A course in language teaching: practice and theory*. Cambridge: Cambridge U Press, 2012. Print.