The Relationship between Vocabulary Size and Reading Comprehension of ESL Learners

Engku Haliza Engku Ibrahim¹, Isarji Sarudin¹ & Ainon Jariah Muhamad¹

¹ Centre for Languages and Pre-University Academic Development, International Islamic University, Malaysia

Abstract

There are many factors that contribute to one's ability to read effectively. Vocabulary size is one important factor that enhances reading comprehension. The purpose of the study is to examine the relationship between students' reading comprehension skills and their vocabulary size. A total of 129 pre-university students undergoing an intensive English language programme at a public university in Malaysia participated in this study. A correlational analysis was employed to ascertain the relationship between scores in the reading comprehension component of the institutionalised English Proficiency Test (EPT) and the Vocabulary Levels Tests (Nation, 1990). Based on Pearson product moment correlation coefficient, there was a moderate correlation (r=0.641) between scores in the EPT reading comprehension and Vocabulary Levels Tests. The relationship was statistically significant at p<0.01 level. The findings also indicate that all students (100%) were able to fulfil the minimum admission requirements for the reading skill (Band 5.5) in the EPT even though only half of the students (54.3%) reached the mastery level at the 5,000 word level. The findings provide useful insights into the prediction of ESL students' performance in reading and the teaching of vocabulary in the ESL context.

Keywords: reading comprehension, vocabulary size, ESL, English language proficiency

1. Introduction

Reading plays a crucial role in the acquisition of a language, particularly in second and foreign language learning. It is thus essential for educators to ensure that learners acquire adequate vocabulary to be able to read and comprehend academic texts well. However, the connection between reading comprehension ability and vocabulary size is complex and dynamic (Hu & Nation, 2000). Vocabulary knowledge is, therefore, a vital language learning component that has to be considered in enhancing reading comprehension, in addition to knowledge of English grammar and background knowledge. Laufer and Sim (1985) claim that in acquiring a foreign language, a learner needs to have sufficient vocabulary knowledge, subject matter knowledge and syntactic structure. In fact, Curtis (1987) claims that students’ ability to acquire new knowledge could be affected if they have low vocabulary knowledge. Therefore, it is crucial to know what it takes for a learner to understand what he/she is reading specifically when challenged with reading texts of academic nature. For this reason, this study attempts to investigate the relationship between reading comprehension of academic texts and vocabulary size of learners.

2. Literature Review

There are essentially many aspects that should be considered. In addition to the kinds of text read, the two important considerations are the density of unknown words that can be tolerated for comprehension to take place and the amount of vocabulary size that a learner needs to have.

The first aspect to be considered is how many unknown words a reader can tolerate for adequate comprehension of fiction texts to take place. Hu and Nation (2000) measured reading comprehension using multiple choice questions and a cued recall written test. A total of 66 participants who had achieved mastery level at the 2,000 word level were required to read fiction texts written using words in the first 2,000 words of English. They experimented by replacing non-words in their texts at the rate of 100%, 95%, 90% and 80%. The participants were randomly assigned to take the tests after reading the four manipulated texts with non-words. The results of
their study indicated that learners need to understand at least 98% of the text read. This can be translated to a
density of 1 in 50 unknown words. The results of Hu and Nation’s (2000) study support the findings of West
(1953) and Hirsch and Nation (1992). This also concurs with Bonk’s (2000) findings that learners who knew less
than 80% of the vocabulary in a text were frequently found to have poor comprehension.

In fact Schmitt, Xiang Ying and Grabe (2011) conducted a study on 661 participants and found that there is a
linear relationship between vocabulary growth and learner’s text comprehension. Their study also highlights the
importance of the threshold level for text comprehension. This is supported by Bonk (2000) who mentioned that
learners with levels of vocabulary familiarity of less than 75% seldom had good comprehension. These research
findings provide the empirical evidence that higher vocabulary knowledge leads to higher text comprehension.

The second aspect to be considered is the amount of vocabulary needed by learners to be able to read authentic
texts without assistance and with full comprehension. Earlier studies of Schmitt (2000) and Laufer (1992)
indicated that a mastery of the most frequent 2,000 words is essential. Milton and Hopkins (2006) claim that
learners would require a vocabulary of around 4,500 – 5000 word families to be able to cope with the highest
level (C2) on the Common European Framework of Reference (CEFR) reading descriptor (an equivalent of Band
7.5 – Band 9 of the IELTS or a score of 590 – 677 in the paper-based TOEFL), though there are criticisms as to
whether learners would be able to achieve the skills listed in the descriptors. However, studies of Hu and Nation
(2000) and Nation (2006) indicated an even higher level of vocabulary is needed, that is, an estimate of 8,000 –
9,000 words for learners to be able to read texts like novels and newspapers without the assistance of a
dictionary or any other source outside the text. Needless to mention, a lower vocabulary size is needed for
learners to be able to read graded texts where texts were written specifically for language learners at the various
vocabulary levels. Thus, it can be concluded that while some learners could cope with a small size of vocabulary,
a large size of around 8,000 – 9,000 words is required for them to be able to read a variety of authentic texts
(Schmitt, 2008).

3. Research Questions

This study was undertaken to investigate the relationship between reading comprehension of academic texts and
vocabulary size of pre-sessional students of the International Islamic University Malaysia (IIUM). Specifically,
this study was carried out to address the following research questions:

1) How do IIUM pre-sessional students perform in the reading comprehension test?
2) How do IIUM pre-sessional students perform in the vocabulary levels test?
3) What is the relationship between reading comprehension and vocabulary size of IIUM pre-sessional students?

3. Methodology

3.1 Setting

The International Islamic University Malaysia is an English medium university. All new intake students are
required to fulfil a minimum English language proficiency of EPT Band 6 (the equivalent of TOEFL 550 or
IELTS Band 6) before they are allowed to undertake credit-bearing faculty courses. Students who do not meet
the minimum language requirement would be placed in one of the 6 levels of the pre-sessional intensive English
language programme offered by the Centre for Languages and Pre-University Academic Programme.

3.2 Participants

The participants in this study were 129 post-secondary students studying English at the Centre for Languages
and Pre-University Academic Programme, International Islamic University Malaysia (IIUM). They were
students from four levels of the pre-sessional intensive English language programme offered by the Centre. The
distribution of the students according to their respective levels is presented in Table 1. A total of 28 students were
from LEM 0320 (Level 1), 28 from LEM 0420 (Level 2), 40 from LEM 0520 (Level 3), and 33 from LEM 0620
(Level 4).
Table 1. Profile of students according to levels of proficiency

<table>
<thead>
<tr>
<th>Course Code</th>
<th>No. of Students</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEM 0320</td>
<td>28</td>
<td>21.7</td>
</tr>
<tr>
<td>LEM 0420</td>
<td>28</td>
<td>21.7</td>
</tr>
<tr>
<td>LEM 0520</td>
<td>40</td>
<td>31.0</td>
</tr>
<tr>
<td>LEM 0620</td>
<td>33</td>
<td>25.6</td>
</tr>
<tr>
<td>TOTAL</td>
<td>129</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Of the 129 students, 75 (58.1%) were females and 54 (41.9%) were males (Table 2). The statistics below reflect the female-male ratio of students at the IIUM.

Table 2. Profile of students according to gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>No. of Students</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>54</td>
<td>41.9</td>
</tr>
<tr>
<td>Female</td>
<td>75</td>
<td>58.1</td>
</tr>
<tr>
<td>Total</td>
<td>129</td>
<td>100.0</td>
</tr>
</tbody>
</table>

3.3 Instrument

Two instruments employed in this study are the vocabulary tests and the reading comprehension test. The Vocabulary Levels Tests (VLT) Version 2 (Schmitt, Schmitt, & Clapham 2001) were used to assess the pre-sessional students’ vocabulary levels. This study adopts the above test as these are the tests that have been widely used to measure L2 students’ vocabulary size. They have been tested for reliability for the 2,000 Word Level Test (Cronbach Alpha of 0.922), 3,000 Word Level Test (Cronbach Alpha 0.927), and the 5,000 Word Level Test (Cronbach Alpha 0.927) (Schmitt, Schmitt, & Clapham, 2001). For the purpose of this study, vocabulary size refers to the word families as defined by Nation (2001: 8) which consists of “…a headword, its inflected forms, and its closely related derived forms. “ This includes affixes such as -ly, un- and -ness.

The second instrument used is the Reading Comprehension Test of the IIUM’s English Proficiency Test (EPT). The EPT is an institutionalized English language test designed to measure the English language proficiency of ESL students in four language skills; namely, reading, writing, listening and speaking. The reading paper consists of 40 multiple choice questions based on four reading passages. The scores are distributed according to Bands, ranging from Band 1 (lowest) to Band 9 (highest).

3.4 Procedure

The students were briefed on the research procedures and the purpose of the vocabulary tests, which was to find out the extent of their vocabulary knowledge. Consent was obtained before the Vocabulary Levels Tests and the reading comprehension tests were administered to the 129 pre-sessional students. They were instructed to complete every item, and not to leave any blanks. All 129 students were present to complete the VLT tests. The same 129 students sat for the EPT reading comprehension test one week later. Tokens of appreciation were given after the completion of both the VLT and the reading comprehension test.

The second procedure was a correlational analysis. The main aim was to ascertain the relationship between scores in the reading comprehension of the in-house English Proficiency Test (EPT) and Vocabulary Levels Tests (Nation, 1990) of 129 pre-sessional students.

4. Findings and Discussion

Findings and discussion are presented based on the three research questions formulated for the purpose of this study.

Research question 1:

How do IIUM pre-sessional students perform in the reading comprehension test?

Table 3 shows the students’ scores in the reading comprehension test of the EPT. Of the 129 students who sat for
the reading comprehension test, 2 (1.6%) achieved Band 5.5, while 7 (5.4%) of the students managed to get Band 9 (the highest Band). Band 5.5 is the minimum English language admission requirement for reading. Thus, all 129 students (100%) were able to fulfil the minimum admission requirements of Band 5.5 in reading comprehension. It is interesting to note that half of the students (49.6%) were able to achieve Band 8. The findings indicate that as far as reading comprehension is concerned, the students not only fulfil the minimum admission requirements of Band 5.5, but also perform 2.5 bands higher. The findings also highlight the fact that even though the students are at the lower elementary level (Level One) to intermediate level (Level Four) of English language proficiency based on the English Proficiency Test (EPT), their reading ability and, to a certain extent, vocabulary size are quite advanced when compared to other language skills such as speaking, writing, and listening.

Table 3. Reading comprehension test scores of pre-sessional IIUM students

<table>
<thead>
<tr>
<th>Band</th>
<th>No. of Students</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>6.0</td>
<td>20</td>
<td>15.5</td>
</tr>
<tr>
<td>7.0</td>
<td>36</td>
<td>27.9</td>
</tr>
<tr>
<td>8.0</td>
<td>64</td>
<td>49.6</td>
</tr>
<tr>
<td>9.0</td>
<td>7</td>
<td>5.4</td>
</tr>
<tr>
<td>TOTAL</td>
<td>129</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Research question 2:
How do IIUM pre-sessional students perform in the vocabulary levels test?

The results of the Vocabulary Levels Test scores of the pre-sessional students are presented in Table 4. The highest mean score for the vocabulary test was for 2,000 word level (M=26.82; SD=3.886), while the lowest mean score was for 10,000 word level (M=6.97; SD=5.55).

Table 4. Vocabulary Levels Test scores of pre-sessional IIUM students

<table>
<thead>
<tr>
<th>Word Level</th>
<th>No. of Students</th>
<th>Mean (%)</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>129</td>
<td>26.82 (89.17%)</td>
<td>3.886</td>
<td>.342</td>
</tr>
<tr>
<td>3000</td>
<td>129</td>
<td>24.03 (79.87%)</td>
<td>4.484</td>
<td>.395</td>
</tr>
<tr>
<td>5000</td>
<td>129</td>
<td>16.46 (54.30%)</td>
<td>6.512</td>
<td>.573</td>
</tr>
<tr>
<td>10000</td>
<td>129</td>
<td>6.97 (23.0%)</td>
<td>5.615</td>
<td>.494</td>
</tr>
</tbody>
</table>

The findings indicate that the majority (80%) of the students have acquired vocabulary mastery at 2,000 and 3,000 word levels. Vocabulary mastery at 2,000 and 3,000 word levels are assumed to indicate that the students have not reached the necessary vocabulary size to undertake faculty courses. At the same time, more than half (54%) of the students in this study are at the 5,000 vocabulary mastery level, which implies that they have reached sufficient vocabulary level to undertake credit-bearing faculty courses. Findings of this study suggest that there is a need to enhance students’ vocabulary knowledge at the 2,000 and 3,000 word levels.

Research question 3:
What is the relationship between reading comprehension and vocabulary size of IIUM pre-sessional students?
A correlational analysis was conducted to investigate the relationship between students’ reading comprehension skills and vocabulary size (Table 5). There is a positive and upper moderate relationship ($r=.641$) between reading and vocabulary test scores. The relationship is statistically significant at $p<0.01$ level. This indicates that the higher the scores in the reading test, the higher the scores are in the vocabulary test. The $R$ Square and Adjusted $R$ Square values are .411 and .392, respectively. The weighted combination of the predictor variables of vocabulary levels 2000, 3000, 5000, and 10000 explained approximately 40% of the variance of reading test scores. The finding suggests that the remaining variance (60%) in the reading test scores is attributed to factors other than vocabulary size.

Table 5. Model summary

<table>
<thead>
<tr>
<th>Model</th>
<th>$R$</th>
<th>$R$ Square</th>
<th>Adjusted $R$ Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.641(a)</td>
<td>.411</td>
<td>.392</td>
<td>.6648</td>
</tr>
</tbody>
</table>

Predictors: (Constant), Level 2000, Level 3000, Level 5000, Level 10000.

Table 6 reports the analysis of the relationships among students’ scores of reading comprehension and vocabulary levels tests for 2000, 3000, 5000, and 10000 word levels. The findings indicate positive relationships between reading test scores and all vocabulary word levels. There was an upper moderate correlation between reading comprehension scores and the 2,000 vocabulary word level with a correlation coefficient of 0.637. Reading comprehension scores also had a moderate relationship with the 3,000 word level ($r=0.496$) and 5,000 word level ($r=0.414$). The 10,000 word level, however, had a weak relationship with the reading test performance. The correlation coefficient was 0.291. The Pearson product moment correlation coefficient reveals that all vocabulary word levels (2000, 3000, 5000, and 10000) had statistically significant relationship ($P<0.01$) with the reading comprehension scores. The findings indicate that vocabulary 2000 word level has the highest predictor value ($r=0.637$) for the reading comprehension scores.

Table 6. Correlations among reading comprehension and 2000, 3000, 5000, and 10000 vocabulary word levels

<table>
<thead>
<tr>
<th></th>
<th>Reading score</th>
<th>2000 word level</th>
<th>3000 word level</th>
<th>5000 word level</th>
<th>10000 word level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading score</td>
<td>Pearson Correlation</td>
<td>.637(**)</td>
<td>.496(**)</td>
<td>.414(**)</td>
<td>.291(**)</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
</tr>
<tr>
<td>2000 word level</td>
<td>Pearson Correlation</td>
<td>.496(**)</td>
<td>.744(**)</td>
<td>.571(**)</td>
<td>.367(**)</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
</tr>
<tr>
<td>3000 word level</td>
<td>Pearson Correlation</td>
<td>.414(**)</td>
<td>.571(**)</td>
<td>.678(**)</td>
<td>.486(**)</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
</tr>
<tr>
<td>5000 word level</td>
<td>Pearson Correlation</td>
<td>.291(**)</td>
<td>.367(**)</td>
<td>.486(**)</td>
<td>.641(**)</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.001</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.1 level (2-tailed).
5. Discussion

The aim of the study was to examine the relationship between reading comprehension skills and the vocabulary size of ESL pre-sessional students in an intensive English language programme. Based on the students’ performance in the Vocabulary Levels Test (Nation, 1990) and EPT’s reading comprehension test, three key findings emerged. Firstly, there is a positive and upper moderate relationship (r=0.641) between students’ reading comprehension scores and their vocabulary size. The relationship is statistically significant at p<0.01 level. The correlational analysis demonstrates that the higher the scores in the reading comprehension test, the higher are the scores in the vocabulary levels test. This is expected considering both tests are measuring similar construct of English language proficiency; in particular, proficiency in reading. The highest correlation is between reading comprehension and vocabulary mastery at the 2,000 word level (r=0.637; p<0.01).

The second key finding relates to students’ performance in the EPT’s reading comprehension test. All students (100%) managed to achieve the minimum faculty admissions requirement of Band 5.5 for reading. Furthermore, half (50%) of the students were able to achieve Band 8. This is an important finding because even though these students are at different levels of the overall English language proficiency (elementary to intermediate), all of them have essentially fulfilled the admission requirement for the reading component of the EPT. This finding concurs with other findings on Malaysia students that they perform better in reading as compared to other language skills such as writing, listening, and speaking (Sarudin, Zubairi, & Nordin, 2008; Engku Ibrahim, Othman, Sarudin, & Muhamad, 2013; Sarudin, Zubairi, Nordin, & Omar, 2008). This is perhaps attributed to the fact that students are more exposed to reading than writing, speaking or listening skill. This is particularly true given that in learning English as a second or foreign language, reading is an important skill that needs to be mastered in order to gain knowledge (Anderson, 1982). In the Malaysia context, English is taught as a second language. Thus, the ability to read in English provides the needed support for learners to be proficient in English for the reason that English can be acquired through reading (Fatimah & Vishalache, 2006). Ultimately, reading is a vital language skill for success beyond academic activities. The findings of a study conducted by Kirsch & Guthrie (1984) underscore the important contribution of reading towards career development and success, and the capacity to respond to new challenges.

The third key finding of this study relates to students’ performance in the vocabulary levels test. About 90 percent of the students managed to reach the 2,000 vocabulary mastery level, while 54.3 percent achieved the 5,000 vocabulary mastery level, and 23 percent achieved the 10,000 vocabulary mastery level. The 5,000 vocabulary mastery level is indicative of the vocabulary size expected of partial college level work, while the 10,000 vocabulary mastery level assumes that the students have reached the vocabulary size for college level work. This is in contrast to the performance of students in the EPT’s reading comprehension test, whereby all (100%) managed to fulfil the minimum admission requirement of Band 5.5, and are eligible to undertake faculty courses. What are possible explanations for the variations in students’ performance? For this group of students, it is reasonable to predict that they may embark on initial college level work once the reach the 2,000 vocabulary mastery level instead of 5,000 or 10,000 vocabulary mastery levels. This conclusion is supported by the results of students’ performance in the 2,000 vocabulary mastery level (90%), EPT’s reading comprehension test (100%), and the correlation (r=0.637; p<0.01) between the two variables as compared to all other variables.

At the same time, a brief analysis of the format of both reading comprehension and vocabulary levels tests may also explain the variations in students’ performance. In the vocabulary levels test, students are required to match a given word with its correct definition from a list of words and a list of definitions without the benefit of contextual clues. This is contrary to the interactionalist approach (Read & Chapelle, 2001) of what a vocabulary test should be as test takers should be able to fall back on contexts in trying to make sense of the vocabulary; a possible explanation as to the difference of performance of our students in the Vocabulary Levels Test and the reading comprehension test of the EPT. In the reading comprehension test students answer reading comprehension questions based on a context, in particular, a reading paragraph. Students are able to employ relevant reading subskills such as skimming, scanning, inferencing, predicting, and contextual clues, among others, in order to answer the reading comprehension questions correctly instead of relying on a collection of definitions to choose from. The reading comprehension test reflects the tasks and contexts required of students to perform in an academic setting. It is also reasonable to assume that for this population, the reading comprehension test of the EPT could possibly be a better predictor of students’ reading ability as compared to Vocabulary Levels Test (Nation, 1990).

6. Conclusion

Although the nature of the research sample and the use of correlation statistics restrict the generalizability of the
findings in terms of cause and effect analysis, some general pedagogical implications could be drawn for colleges that share similar demographic features. Specifically, it is essential to highlight the role of teachers to make available words at the 2,000 level so that students can be exposed to these words in their daily reading or entertainment literacy encounters. Given these students’ vocabulary size, it is recommended that they continue to develop their knowledge of high-frequency words at the 5,000 and 10,000 word levels and, at the same time, expand their knowledge of low-frequency words. Furthermore, teachers need to play a more active role in creating awareness of the importance of vocabulary related activities in building students’ vocabulary size.

Teachers should also encourage students to engage in extracurricular extensive reading activities (e.g. Zhang, 2001b, 2003), as there is some cumulative evidence indicating the benefit of extensive reading in helping learners to enhance vocabulary size and reading abilities (Day & Bamford, 1998; Krashen, 2004; Nation, 2001). At the same time, Hunt and Beglar (2005) propose a systematic framework for lexical development in order to speed up lexical development; an aspect that is particularly true for the context of this research as learners have very limited time to master the English language prior to pursuing their respective degrees. Likewise, it is important for learners to be exposed to reading subskills of predicting and guessing from contexts in order to compensate for the low vocabulary size. Needless to mention, learners also need to realise that vocabulary acquisition is an important life-long skill and that they need to be able to acquire more vocabulary independently throughout their academic life and beyond. Ultimately, critical reading strategies, which focus on evaluating and appraising the quality, value and truthfulness of the reading, may be gradually introduced to enhance not only students’ reading, but also critical thinking skills.

References


Copyrights
Copyright for this article is retained by the author(s), with first publication rights granted to the journal.
This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).