

## CHAPTER TWO : LITERATURE REVIEW

*“It’s important to have good vocabulary. If I had known the difference between the words ‘antidote’ and ‘anecdote’, one of my good friends would still be living.”*

–John McDowell (1992), a comedian, on ‘The Importance of Vocabulary’.

The importance of vocabulary is aptly illustrated in John McDowell’s quote. From a drowning man shouting for ‘Help!’ to a lawyer trying to defend his innocent client, vocabulary is our most used and best communication tool. Vocabulary is used in play, in fight, in dream, to buy groceries or build a bridge. People express emotions, direct, teach and persuade, for a variety of reasons in a variety of situations to a variety of audience, all of which uses one common tool - vocabulary.

Human beings are social animals. We are very much reliant on using language to connect with others. The building blocks of any language would be its words, or vocabulary. A person may know all the intricacies of grammar and structure of the language, but yet, may fail to get his message across without the appropriate vocabulary. “Lack of grammatical knowledge sometimes impedes successful transmission of meaning; however, absence of vocabulary often impedes the transmission of meaning completely” (Wilkins, 1972, cited in Barcroft, 2004, p.201).

This is more so for second or foreign language learners. To express an idea or communicate effectively can be the most frustrating experience in L2 communication when “one is not being able to find the words that one needs” (Wallace, 1982, cited in Nurnia 2000, p.3). L2 students who sit for exams may be terribly frustrated when a crucial word cannot be found or recalled.

Research has also shown that students attribute vocabulary knowledge highly in their ability to function in the second language (James, 1996, cited in Barcroft, 2004). They perceive that having a big vocabulary bank would make them more proficient. Through personal experience, some students in Malaysia have been known to resort to memorizing words from the dictionary in the hope of scoring better in their English language tests.

In order to explore vocabulary in greater depth, this chapter will start off the discussion by defining vocabulary and looking at its nature. It will then look into the teaching and learning processes of vocabulary acquisition before moving on to a discussion on vocabulary size, selection and how it is measured and tested. The discussion will then take on more specific issues by examining past researches and studies on vocabulary in relation to its implication on language proficiency and how the ecological factor plays a role.

## **2.1 Vocabulary**

From experience and discussion with colleagues, a language teacher would term vocabulary as ‘difficult words to be taught’. A science teacher would term vocabulary as ‘terminology or jargon used in the science’. A layman would term vocabulary as simply, ‘words’.

According to linguists, vocabulary is an umbrella term that encompasses many entities. First of all, there is the ‘lexeme’. A lexeme is defined as “an item that functions as a single meaning unit, regardless of the number of words it contains” (Schmitt, 2000, p.2). ‘Fun’ is a lexeme, and so are ‘make merry’ and ‘have a ball’. Secondly, there is the ‘word family’. This includes the base or root form of words as well as its inflections and derivatives (Schmitt, 2000). For example, ‘respect’ is a root word. Its inflections include

‘respects’, ‘respected’ and ‘respecting’, where affixes are added for the purpose of grammar. The root word and its inflections are also termed as ‘lemma’. Its derivatives include ‘respectable’ and ‘respectful’, where affixes are added for the purpose of changing the word class.

In this study, the word ‘vocabulary’ would simply mean ‘words’. They will be used interchangeably for the ease of discussion unless stated otherwise. To examine vocabulary further, a discussion on the nature of vocabulary may be useful.

## **2.2 The Nature of Vocabulary Knowledge**

Knowing a word truly may not be as easy as it seems. Experts in the field have studied the nature of vocabulary knowledge and have discovered a few of its characteristics.

First of all, vocabulary knowledge is said to be ‘incremental’ in nature (Schmitt, 2000). It is not acquired or learned instantly. To know a word requires time, repeated exposures and use. Knowledge of a word is widely categorized as the receptive knowledge (the ability to understand and recognize) and the productive knowledge (the ability to use). The general assumption is that a word is first recognized and understood (receptive), before it is used (productive). This linear progression suggests that vocabulary knowledge is sequential in nature. However, Schmitt (2000) has suggested otherwise where a person is able to use a particular word in speech, before being able to recognize it in a written text. Hence, the incremental nature of vocabulary may not happen sequentially in a linear progression.

However, the incremental nature of vocabulary knowledge may be explained and quantified by its degree of mastery (Schmitt, 2000). To illustrate, receptive knowledge of a word involves the recognition of it in a spoken situation and in a written text. It also

involves understanding the meaning of the word used in different situations, for example in a proverb or in the court of justice. Hence, to have a perfect receptive knowledge of the word, a person has to ‘master’ all these knowledge. This includes the mastery of both receptive and productive knowledge. Nation (1990, cited in Schmitt, 2000, p.5) suggests a complete mastery knowledge list which includes the knowledge of its “meaning(s), written form, spoken form, grammatical behaviour, collocations, register, associations and frequency”. In other words, Lehr, Osborn, and Hiebert, (2004, p.4) term them as ‘multidimensional knowledge’ (multiple meanings and functions) and ‘interrelated knowledge’ (relationships or connections).

The idea of vocabulary mastery then puts receptive and productive knowledge on a multi-level continuum of knowledge, from zero knowledge to full mastery. This may suggest that both receptive and productive knowledge is measurable using a specialized yardstick. Experts have since developed the yardstick they call ‘vocabulary test’. The test comes in many forms for many different purposes and results have generally shown that receptive vocabulary is usually greater than productive vocabulary and that the size of one can predict the size of the other. In short, it can be said to be ‘predictable’.

Therefore, apart from being incremental in nature, vocabulary knowledge can also be said to be measurable, multidimensional, interrelated and predictable. With this understanding of the nature of vocabulary knowledge, it would lead to a discussion on its teaching and learning processes.

### **2.3 Vocabulary Teaching and Learning**

As teaching usually imitates the learning process, it may be best if the learning process is explored first in this section. The discussion on the learning process will then provide a foundation to the discussion for better strategies and methodologies in teaching.

Previous researchers (Nation and Waring, 1997, cited in Schmitt, 2000) have estimated that a native speaker acquires almost 1000 word families per year. It would be a fallacy to say that a person learned all of them in a year only through formal schooling. As vocabulary mastery is said to be incremental, measurable, multidimensional, interrelated and predictable, as discussed earlier, the learning process of this large number of words will take place not only through formal instruction, but also through informal situations. A better categorization for the learning situations would be explicit (formal) and incidental (informal) learning.

L1 users of language are generally said to gain vocabulary through incidental learning. They are exposed to the language in vast amounts from the day they were born and also have vast opportunity to use and practice in the language. L2 users have also been generally assumed to gain most from incidental exposure to the language, as the L2 language environment is generally believed to be rich enough, albeit a more active explicit learning in school, compared to L1 users. However, Schmitt (2000, p.121) argued that “one of the most common problems” faced by them is the lack of exposure for incidental learning to contribute to their vocabulary gains. Other researchers (Long, 1997; Jiang, 2000; Lardiere, 1998, cited in Chen, 2009) have also noted this limitation to L2 vocabulary gains, whereby they have suggested that even in situations where ample opportunities for practice as well as contextualized input (and motivation) are available, most L2 learners never achieve target-like proficiency and their interlanguage is prone to “fossilization”. This may indicate that L2 learners’ vocabulary development may be ‘stunted’ to a certain extent.

Hence, the National Reading Panel (2000, cited in Lehr et.al., 2004, p.4) of the United States, thus, concluded that “no one single instructional method is sufficient for optimal vocabulary learning”. Explicit and incidental learning are complementary. A five

hour per week language lesson in school or language course may not be adequate to equip a person to have enough vocabulary knowledge to function, especially in an L2 situation. His or her exposure to words learnt explicitly can be reinforced with words heard, read and used in daily dealings (incidental exposure) and vice versa.

This symbiotic relationship is said to bring about optimal benefits when the opportunities for exposure, use, manipulation and thinking are present and abundant (Schmitt, 2000). Hence, in vocabulary teaching, Sökmen (1997, cited in Schmitt, 2000, p.146-147) highlights a number of “key principles” which include:

- a) building a large sight vocabulary
- b) integrating new words with old
- c) providing a number of encounters with a word
- d) promoting a deep level of processing
- e) facilitating imaging
- f) making new words ‘real’ by connecting them to the student’s world
- g) using a variety of techniques
- h) encouraging independent learning strategies

All these principles are echoed in the eight teaching conditions set by the National Reading Panel of America (2000, cited in Lehr et.al., 2004, p.5):

1. There is a need for direct instruction of vocabulary items required for a specific text.
2. Repetition and multiple exposures to vocabulary items are important. Students should be given items that will be likely to appear in many contexts.
3. Learning in rich contexts is valuable for vocabulary learning. Vocabulary words should be those that the learner will find useful in many contexts. When vocabulary items are derived from content learning materials, the learner will be better equipped to deal with specific reading matter in content areas.
4. Vocabulary tasks should be restructured as necessary. It is important to be certain that students fully understand what is asked of them in the context of reading, rather than focusing only on the words to be learned. Restructuring seems to be most effective for low achieving or at-risk students
5. Vocabulary learning is effective when it entails active engagement in learning tasks.
6. Computer technology can be used effectively to help teach vocabulary.

7. Vocabulary can be acquired through incidental learning. Much of a student's vocabulary will have to be learned in the course of doing things other than explicit vocabulary learning. Repetition, richness of context, and motivation may also add to the efficacy of incidental learning of vocabulary.
8. Dependence on a single vocabulary instruction method will not result in optimal learning. A variety of methods was used effectively with emphasis on multimedia aspects of learning, richness of context in which words are to be learned, and the number of exposures to words that learners receive.

Various methodologies, strategies and programmes have also been created according to these guidelines. Some of these include intensive and extensive reading, guided readers, reading aloud, guessing words from context, mnemonics, word games, the Big Book Approach, Computer-Aided Language Learning (CALL) programmes, language softwares and specialized dictionaries.

In the case of L2 learners, some of the conditions mentioned earlier may be limited due to ecological, socioeconomic, attitude or time factors. It is suggested that a concerted effort is needed by the individual himself or herself, parents, family members and teachers to enhance the learning experience (Lehr et.al., 2004).

The types of teaching and learning methods employed have also proven to have effects on the size of vocabulary knowledge. Studies have shown that learners ended up with larger vocabularies when teachers employ/use a variety of teaching and learning strategies. One example to illustrate this is as children progress and begin formal schooling, they gain larger vocabulary size, by consistently "being engaged in interactive teacher-child talk and storybook reading" (Schwanenflugel et.al., 2004, cited in Lehr et.al., 2004, p.6), compared to only reading aloud by itself.

In Webb's (2008) study, he found that the type of teaching and learning methods, namely explicit and incidental learning, has affected the size of the gap between receptive and productive knowledge. In a more explicit teaching and learning situation, as adopted

by EFL learners, the gap between their receptive and productive knowledge is noticeably smaller, compared to the larger gap of L1 and L2 students, who tend to learn more incidentally.

Therefore, different methods in the teaching and learning of vocabulary may result in different vocabulary size. With this in mind, the discussion on vocabulary size proceeds.

## **2.4 Vocabulary Size**

The English language has been reported to have at least 54,000 word families (Goulden et al., 1990, cited in Schmitt, 2000). It is quite impossible to acquire the complete bank of words, even in the case of native speakers. However, native-speaking university graduates are estimated to have mastered around an exceptional 20,000 word families (Schmitt, 2000).

For second language learners, this feat would be almost un-doable, given the same amount of time, as L1 and L2 learners acquire vocabulary differently. L2 learners are at an obvious disadvantage as they have less exposure to the language. Hence, they need an achievable target to function well in the language.

The milestone in vocabulary has been set by categorizing words into frequency levels and specific groups. According to Nation's word levels, the 2000 word families which appear most often are categorized into the high frequency words category. This includes function words and content words. "They account for almost 95% of the running words in a text" (Nation, 2002, p.6). The next frequency levels would be the 3000, 5000 and 10 000 word level, with each level declining in frequency.

The Collins COBUILD Dictionary (2001) also categorizes its entries into bands of frequency. Its entries were derived from the 400 million word Bank of English corpus;



COBUILD provides five frequency bands, from 680 (Band 5), 1720 (Band 4), 3300 (Band 3), 6500 (Band 2) to 14600 (Band 1), with Band 5 containing the most frequent words and Band 1, the least frequent words. It also includes words less frequent than Band 1, which deserved to be mentioned. No band has been provided for these words.

Apart from frequency bands, vocabularies have also been categorized into specific groups. This includes the ‘Academic Word List’. The vocabularies included here are words familiar in academic texts from a variety of subjects, from astronomy to zoology. They are typically not in the first 2000 frequency word level (Nation, 2002). Another group is the ‘Technical Words’. Words grouped here are specific to its particular subject only, for example medicine, music or mathematics. To illustrate, the word ‘scale’ is interpreted differently in all three subjects.

Given these set categories, learners can target their vocabulary sizes accordingly. Schmitt (2000) stated that 2000 of the most frequent word seems to be the most commonly cited initial goal for second language learners. This is because learners who know these word families comprehend about 75% of what they read (Stahl and Nagy, 2006, cited in Bates, 2008), and 90% of what they hear (Nation, 2001, cited in Webb, 2008). When learners master the Academic Word List, containing 570 additional word families, it brings learners to an additional 4% coverage of newspapers and up to 10% coverage of academic texts (Nation, 2002).

Having discussed this, the next logical step would be to explore the criteria to select words to be taught and learned.

## **2.5 Vocabulary Selection**

Research from the long history of corpus linguistics has shown that “all words are not created equal” (Nation, 2002, p.6). Some occur much more frequently in our everyday

lives, while some deserve more attention (perhaps for educational purposes), while others are only known by specific groups of people, in their respective professional fields. Some words may also only appear more frequently in conversations and some only in written texts. There are 400 million words in the Bank of English corpus with different degrees of importance for different walks of life.

For the selection of words, Nation (1995, cited in Schmitt, 2000), on the one hand, considers its cost/benefit. Nation (2000) explains that any time spent on words should be in consideration of its benefits. This may suggest that time is considered well spent when investing in high frequency words, which not only provide high coverage but is also manageable in size. Low frequency words however, have a poor coverage and therefore not worth investing time on.

On the other hand, Bates (2008, p.68) urges teachers to adopt “responsible vocabulary word selection”. He argues that textbooks publish word lists which contain rarely used words with minimal usefulness, which he termed as ‘50-cent words’. Hence, he calls for teachers to first of all determine students’ entry level vocabulary, as the learning of more difficult words depends on the understanding of more simple words. He cautions against skipping word levels, as it can have serious consequences, which may lead to cumulative effects. Next, he recommends that the target vocabularies should be a level higher than the entry level, in line with Krashen’s (1981) “*i + 1* hypothesis”. The selection process then should be guided by empirical word studies, namely the Actual Word Knowledge Studies (the actual vocabularies of students at a particular grade level), Word Frequency Studies (the frequency of a word in a corpus of millions) and Textbook Word Frequency Studies (the frequency of words appearing in textbooks and literature selections at particular grade levels). He also identified characteristics that make certain vocabulary better for instruction. This includes having connection to literature at the

targeted level, words from word families, cross-curricular words, words sharing cognates with L1, words that frequently appear in tests, words from a theme and words with commonly used roots, prefixes and suffixes.

Hence, for a more cost-effective and responsible vocabulary selection, students' current level of vocabulary would have to be measured.

## **2.6 Vocabulary Tests**

The best litmus test for vocabulary may be the daily newspaper (Sawyer, n.d.). If one has problems reading through it, then perhaps, one's vocabulary size may be inadequate. However, experts in the field have designed various empirical tests to determine one's vocabulary size. There are two principle methods of vocabulary tests. The first is the controlled test, where formats include matching, cued recall, checklist, assessment scale and word association. The second is the free production test, which takes the form of writing or speaking tasks.

The most well known of all the tests would be Nation's (2000) Vocabulary Levels Test (VLT). It is a controlled test in the form of a matching test. There are five levels of word frequency to be tested: the 2000, 3000, 5000, 10000 and the academic word level. A selection of 60 sample words are chosen from each level and grouped into blocks of six, where only three would be tested. A definition of these words, which are more frequent than the tested words, are given to be matched. The VLT can be a useful receptive knowledge test and Schmitt (2000) commented that it is useful for placement purposes and for diagnosing vocabulary gaps as it provides a reading at five different levels. However, the VLT was also criticized as it provides options of words for test takers to match, hence, providing them a possibility to score correctly even without knowing the word. Morris and Cobb (2004) likened this to 'passive vocabulary test' whereby its

multiple choice format opens doors for candidates to score correctly by purely guessing and even memorizing answers, without having knowledge for something as basic as spelling.

The Productive Vocabulary Levels Test (PVLТ) developed by Laufer and Nation (1999, cited in Webb, 2008) complements the VLT. It tests the productive knowledge using the cued recall test format at five levels of word frequency. Test takers have to complete a word in a sentence guided by the first letter of the missing word. Experts were concerned that it may lack validity as it may actually be testing receptive knowledge, as Morton's (1979, cited in Webb, 2008) research and Melka's (1997) statement show that partial information is sufficient to recognize a word. Webb (2008, p.80) further argues that "there is a disparity between the receptive and productive tests" whereby the latter is more linguistically demanding as it involves more aspects of knowledge including spelling and grammar. Furthermore, the difference in format, where the former uses a recognition format and the latter, a recall, would not bring about a fair comparison. On top of that, only selected words are tested from a pool of more than 2000 words for each level and this may serve an insufficient basis for assessment.

An example of the checklist test is Meara's (1992, cited in Schmitt, 2000, p.175) EFL Vocabulary Tests. Test takers would be given a list of words as well as nonwords and are required to put a tick if they know the word. A formula is used to calculate whether the testee has overestimated his or her vocabulary knowledge, by taking into account the number of nonwords ticked. This can be a simple test conducted for the purpose of placement (Schmitt, 2000). This checklist format has also been questioned as test takers may overestimate their vocabulary knowledge and tick more words than they actually know. In contrast, the test takers may also tick many nonwords and render it unreliable (Schmitt, 2000).

Paribakht and Wesche (1993, cited in Schmitt, 2000) designed a test based on self-assessment scale, called the Vocabulary Knowledge Scale (VKS). It combines student self-report with production to ensure student know how to use the words. The scale of knowledge starts from not having seen the word before, to knowing the meaning of the word. Test takers have to either construct a sentence or provide a synonym or translation. It has gained popularity in recent years as it is said to be more motivating. Schmitt (2000) commented that it allows test takers to focus on what they know instead of what they do not know. It also provides the opportunity for them to show their partial or full knowledge. As with other tests, the VKS also has its limitations. Schmitt, points out that the scale mixes both receptive and productive knowledge, hence making it difficult to interpret. It is also debatable since the number and increment of levels of the scales have been empirically tested.

The word association test can be exemplified by Read's (1993, 1998, cited in Schmitt, 2000) Word Associates Test and Vives Boix's (1995, cited in Schmitt, 2000) Association Vocabulary Test. The latter requires test takers to choose four out of eight word options associated with the target word. It is one of the first attempts to measure associative, collocational and conceptual knowledge in two categories; the paradigmatic and syntagmatic. Read himself (1993, 1998, cited in Schmitt, 2000), considers this test unreliable due to the probability of guessing. In Boix's test, test takers are required to choose one unrelated word out of a three-word cluster. This rectified the guessing element and she believes that this may be a good supplement to size tests, as it measures the degree of lexical organization (Schmitt, 2000).

The free production of writing or speaking tasks can be seen in Laufer and Nation's (1995, cited in Meara and Fitzpatrick, 2000) Lexical Frequency Profile (LFP). The written or spoken words of test takers are analysed using a computer programme to

ascertain lexical density (density of content words), variation (range of words) and sophistication (use of low frequency words). Some researchers are in favour of the profiler as it allows access to additional information derived from words used in a context, for example collocations, type-token ratios and spelling (Morris and Cobb, 2004). However, others have said that it is too context-specific and may not provide the opportunity for test takers to ‘display’ their vocabulary knowledge to the maximum. To gauge the actual vocabulary size, a large amount of data would be needed, which Meara and Fitzpatrick (2000) account for as being not cost-effective.

### **2.6.1 Improved Tests**

In light of these issues, studies were conducted to determine more suitable tests. A few considerations gleaned from the previous review, would include having an equivalent test and scoring format, limiting probabilities of guessing, controlling over-estimation, providing opportunities for production of low frequency words and being cost-effective.

Webb (2008), in his study of EFL students in Japan, found that translation test provides “a more accurate measurement” (p.92) as they have an equivalent test format for both receptive and productive vocabulary. The tests measured vocabulary sizes at three frequency levels; 1900, 3400 and 6600, based on COBUILD dictionary’s frequency information. 60 words were selected for each level, taking into consideration the degree of overlap in L1 and L2 meanings to avoid eliciting varied responses in the translation. Webb’s receptive test requires test takers to translate L2 words into L1 words and vice versa. This test was able to measure test takers’ knowledge in terms of meaning and form from its scoring method; sensitive scoring (meaning) and strict scoring (meaning and form). Webb cautions the use of translation tests in an ESL situation as grading it may prove to be difficult as test takers may have different L1s and the target words need to be

selected carefully and extensively pilot tested to ensure test takers are able to respond to them.

Meara and Fitzpatrick (2000), in their attempt to seek an improved L2 productive vocabulary test, devised an instrument, Lex30. Lex30 tries out the word association format, where 30 stimulus words prompt EFL test takers one at a time, to produce responses in 30 seconds for each stimulus. The stimulus words were selected based on the criteria that they are highly frequent words, do not elicit a single dominant primary response and generate responses which are not common words. A standard yes/no test was also administered to test receptive knowledge. The responses for Lex30 were analysed by lemmatizing them before categorizing them according to frequency level. It has proven to show a high level of internal consistency and some potential to be a diagnostic tool. It can also be easily administered and can be used as part of a larger test battery. It has also displayed the ability to generate a wide variety of responses including uncommon words. Together with a lenient scoring method where subjects are given credit at every possible opportunity, students whose experience of words is influenced by special circumstances or experience are not penalised. Their responses enjoy more freedom for credit compared to controlled tests.

All of the mentioned tests have been designed and developed to study different aspects of vocabulary. A common result from most tests is that receptive vocabulary is found to be larger than productive vocabulary (Schmitt, 2000, and Webb, 2008). Studies have also found that the size of receptive vocabulary predicts the size of productive vocabulary (Meara & Fitzpatrick, 2000). However, Nation cautioned test developers against overgeneralizing while interpreting test results (Coxhead, 2005). This is because the test results may be influenced by the type of test, test takers' background, experience and attitude towards the test.

A discussion on the implications of vocabulary would explore these factors.

## **2.7 Vocabulary and Its Implications**

Many factors have been proven to have an influence on and be influenced by vocabulary. In Malaysia, urban and rural folks seem to have differing standards of English proficiency. This can be observed in the everyday choice of language among the people in public, where English is more frequently heard among urban folks compared to rural folks, where it may not be used at all. This disparity is also apparent in the public English examination results in Malaysia, where urban students have been outperforming rural students. Hence, the following discussion will focus on these two factors, namely the ecological factor and language proficiency, and explore how vocabulary implicates them.

### **2.7.1 The Ecological Factor**

Urban and rural folks enjoy different privileges. While people staying in the rural areas enjoy a less hectic lifestyle with cleaner environment, the urbanites enjoy the vast opportunities of modern development. This ecological factor has long played a role in people's lives, even in the education field.

In terms of academic results, urban students have been outperforming rural students. Studies around the world have found sub-standard achievement to be a pattern in rural areas (McCleery, 1979, cited in Eddington and Koehler, 1987; Easton and Ellerbruch, 1985, cited in Eddington and Koehler, 1987). This is also reflected in the public examination results in Malaysia, including the English language subject.

Lok (2007), the secretary-general of the National Union of Teachers' Profession of Malaysia (NUTP), in her address at the Malaysian 11<sup>th</sup> Education Summit, noted that this pattern is true and still persists because rural schools are at a disadvantage in terms of



resources, aids and attitude of stakeholders. The common problems highlighted are the lack of infrastructure, electricity, transportation and teachers, unfair distribution of aid, non-standardized pre-school curriculum, social problems, socioeconomic status and limited education opportunities and access to ICT.

To examine this issue in terms of language and vocabulary, Rosli Talif and Edwin (1990) conducted a study on the achievement and proficiency level in English as a second language among learners in selected rural and urban schools in Malaysia. Their findings proved that urban students are indeed performing better than their counterparts overall in all the test components of the proficiency test. In terms of vocabulary, urban students scored 84% on average as compared to only 65% for rural students. An analysis of their English language SRP results also showed that the bell curve for urban students swayed towards the distinction level. In comparison, the curve inclined towards the credit and pass level for rural students. They concluded that rural schools contributed heavily towards the high failure rate in the subject in the 1980s.

The Ministry of Education of Malaysia is well aware of this problem and has put in effort to bridge the gap. Through its Education Development Blueprint (PIPP), it has not only pledged to improve the infrastructure in rural schools, but also increase the amount of exposure to English instruction in classrooms by using English in the teaching of Mathematics and Science (ETEMS) throughout the nation (Ministry of Education Malaysia, 2008). In 2012, the move will be revised again where more teaching periods will be assigned to the teaching of English from the current 5 periods, the equivalent of around 3 and a half hours, per week.

As the ETEMS effort has been on-going for almost eight years, the passing rate for English has improved from 282,471 passes in 2008 to 302,501 in 2009, as reported by the Education Director-General (Azreen Hani and Tamboo, 2010). This has also in turn

narrowed the gap between urban and rural students' achievement. Nevertheless, the gap still exists.

This may be due to challenges hidden under the umbrella of the ecological factor. This may include the socio-economic status (SES) of the family. It is said that urban families have higher SES compared to rural families. Studies which have been conducted have found a strong relation between vocabulary knowledge and socioeconomic status (SES). The home environment of rural and urban students may contribute substantially to learners' vocabulary knowledge (Baker et.al. 1995).

Studies in general have found that children in urban areas, from higher SES homes, are at an advantage, where they engage in more interactive discussions with their parents compared to rural children in lower SES homes (Lehr et.al., 2004). Hence, children with plenty of exposure and opportunity to use the language have a better headstart than those receiving less experience, and all these happen before a child enters formal schooling, or even kindergarten.

In a particular study by Hart and Risley (1995, cited in Lehr et.al., 2004, p.6), they found that “three-year-olds in higher SES families had vocabularies as much as five times larger than children in lower SES families”. Researchers have also revealed that when children enter kindergarten, higher SES children have “almost twice the usable vocabulary as low SES children” (Graves, Brunetti and Slater, 1982, cited in Berne and Blachowicz, 2008, p.314). Moats (2001, cited in Lehr et.al., 2004, p.5) termed this gap as ‘word poverty’. The Peabody Picture Vocabulary Test conducted by the Government of Canada (2006) mentioned earlier, also stated that urban children performed better than rural children, indicated by higher development in the language skills and having a larger vocabulary size. The same can be said for L2 learners. Research has shown that the type

of elements that benefit L1 learners also benefit L2 learners (National Literacy Panel, August 2004, cited in Lehr et.al., 2004).

In addition, a review of research by Sirin (2005) noted that family SES is one of the strongest correlates of academic performance, including verbal achievement. This is so because family SES determines the neighbourhood or location (rural or urban) in which they live or go to school as well as the availability of educational resources, for example books, television and the internet. Hence, this concurs with the earlier mentioned Lok's (2007) address on the challenges regarding rural schools.

Rural and urban learners' language achievement gap has also been found to be due to motivational factors. According to another review of literature by Jianzhong Xu (2009), studies have indicated that rural students tend to have lower educational aspirations compared to their urban counterpart. An example quoted by Jianzhong Xu was Hu's (2003) study, which found that the aspiration for rural students to go for tertiary education is lower than that of urban students. The finding was also supported by the smaller enrolment of rural students in postsecondary institutions.

This is also echoed in the voices of rural Malaysian students given the history of the English language in Malaysia. Gaudart (1987) reported that English was only taught in urban schools when Malaya was under the British rule. Only basic Malay was taught in rural areas. Hence, being deprived of English education meant that rural students were prevented from attaining higher status employment. This could have caused rural students to be content and lack motivation to study English.

In summary, this section has discussed how the ecological factor affects students' language and vocabulary gains. The following section will delve deeper into the relationship between language proficiency and vocabulary.

### **2.7.2 Language Proficiency**

Recently, the hype regarding vocabulary is again beginning to rise and it is said to be getting ‘hotter’ (Cassidy and Cassidy, 2005/2006, cited in Berne and Blachowicz, 2008, p.314). This is due to the fact that it has shown to play a major contributing role in the proficiency of language skills. Its role is noted to be so significant that it has been given the honour to be legislated as one of the “foundational pillars” of the curriculum by the Reading First programme in the United States (Berne and Blachowicz, 2008).

Reading proficiency would perhaps have the most proven relationship with vocabulary. Researchers have consistently noted “a strong and persistent link between vocabulary instruction and comprehension based tasks” (Berne and Blachowicz, 2008, p.315). Decoding skills, fluency skills and comprehension skills all draw upon a known bank of words. For example, repeated reading, even when it has been proven to improve reading fluency, will fall short if most of the words in the text are unknown by the students. Teachers who use cue words to encourage vocabulary activities will fail, if the cue words are unknown to the student.

Another study by Chall and Snow (1988) explored factors contributing to what they call the ‘fourth-grade slump’ in terms of poor literacy, in students from low-income families. They found out that for students to achieve the expected literacy level, good and early vocabulary instruction plays a vital role. Students have to have firstly, a good strong start in the primary grades in instruction in word recognition, decoding, and fluency, and in reading widely. This is followed by structured and challenging instruction in reading and vocabulary, opportunities for stimulating, enriched exposure to written materials and opportunities to practice literacy skills. They recommended that techniques for extending literacy and vocabulary instruction should be developed and that students’ progress has to

be monitored with the active participation of teachers and parents, so that the 'slump' does not snowball further.

With a 'slump' in vocabulary, students tend to struggle with words and consequently lose interest in reading and may face frustration as texts become more challenging, whereas students who have sufficient vocabulary knowledge would have less inhibition. This sets in motion Stanovich's (1986, cited in Lehr et.al., 2004, p.2) 'Matthew Effects'. This idea was inspired from the Bible; Matthew Chapter 25 verse 29, where it says that "the rich get richer and the poor get poorer". As good readers read more, they learn more words, whereas poor readers read less and inadvertently, learn fewer words. This idea can also be applied to studies on vocabulary knowledge. Research suggests that the more students are exposed to receptive learning, the more gains they achieve in receptive knowledge than productive knowledge. In contrast, the more practice students get from productive learning, the greater the increase in productive knowledge than receptive knowledge (Webb, 2008).

Apart from reading proficiency, the overall language proficiency has also been proven to rely greatly on vocabulary. In terms of language as an academic subject, vocabulary knowledge is said to be a good predictor of its scores. In Morris and Cobb's (2004) study, their vocabulary profiling of TESL trainees was found to be good predictors of academic performance. It confirms the notion that a better knowledge of academic words (receptive) and a better ability to access a formal academic register (productive), results in better language academic performance that is metacognitively demanding.

Zareva, Schwanenflugel and Nikolova (2005, p.592) also explored the relationship between vocabulary knowledge and the overall language proficiency. Their results showed that "the quantity and quality of vocabulary knowledge (particularly vocabulary

size, word frequency effects, number of associations, and within-group consistency of associative responses) are effective in distinguishing overall language proficiency differences between intermediate and advanced L2 learners”.

Another example of how vocabulary plays a significant role is that the Government of Canada has used vocabulary as a predictor of school readiness for young children entering kindergarten (Government of Canada, 2006). Its Peabody Picture Vocabulary Test measures receptive vocabulary and it shows strong significance in predicting a child’s ability to communicate, learn and integrate into society.

In summary, the discussions and implications of vocabulary from the review of literature warrant further exploration of its intricacies, based on local settings.