A NEW VOCABULARY-BASED READABILITY INDEX FOR THAI UNIVERSITY STUDENTS

Patteera Thienpermpool

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in English Language Studies

Suranaree University of Technology

Academic Year 2009

ดัชนีวัดความยากง่ายของบทอ่านแบบอิงคำศัพท์ใหม่สำหรับนักศึกษาไทยในมหาวิทยาลัย

นางสาวภัทร์ธีรา เทียนเพิ่มพูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต
สาขาภาษาอังกฤษศึกษา
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2552

A NEW VOCABULARY-BASED READABILITY INDEX

FOR THAI UNIVERSITY STUDENTS

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

____

(Dr. Dhirawit Pinyonatthagarn)

Chairperson

____

(Dr. Sarit Srikhao)

Member (Thesis Advisor)

____

(Prof. Dr. Chaiyong Brahmawong)

Member

____

(Dr. Jitpanat Suwanthep)

Member

____

(Dr. Suksan Suppasetseree)

Member

____

(Prof. Dr. Sukit Limpijumnong)

Vice Rector for Academic Affairs

____

(Dr. Peerasak Siriyothin)

Dean of Institute of Social Technology

ภัทร์ธีรา เทียนเพิ่มพูล : ดัชนีวัดความยากง่ายของบทอ่านแบบอิงคำศัพท์ใหม่สำหรับนักศึกษาไทยในระดับมหาวิทยาลัย (A NEW VOCABULARY-BASED READABILITY INDEX FOR THAI UNIVERSITY STUDENTS) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร. เจเรมี วิลเลียม วอร์ด, 351หน้า

การศึกษานี้มีจุดมุ่งหมายเพื่อสร้างดัชนีวัดความยากง่ายของบทอ่านแบบอิงคำศัพท์และตรวจสอบความเที่ยงตรงของดัชนี ผู้เข้าร่วมการวิจัยครั้งนี้มีทั้งหมด 6 กลุ่ม แต่ละกลุ่มมีจำนวนดังนี้ 102, 45, 80, 5, 30 และ 6 ตามลำดับ ข้อมูลที่ได้มาจากกลุ่มแรกใช้ในการสร้างดัชนี ส่วนข้อมูลในกลุ่มที่เหลือใช้เพื่อตรวจสอบความเที่ยงตรงของดัชนี งานวิจัยชิ้นนี้ศึกษาตัวแปรทั้งหมด 4 ตัวแปร: 1) สัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อย 2) คะแนนจากข้อสอบคลังคำศัพท์ 3) เปอร์เซ็นต์ของความรู้คำศัพท์ในแต่ละบทอ่าน 4) ความเข้าใจในการอ่าน เครื่องมือที่ใช้ในการเก็บข้อมูล คือ 1) คอมพิวเตอร์ซอฟต์แวร์ที่เรียกว่าเรนจ์ 2) ข้อสอบแบบรู้/ไม่รู้ 3) แบบรายงานคำที่ไม่รู้จัก 4) ข้อสอบการแปล 5) ข้อสอบวัดความเข้าใจในการอ่าน และ 6) แบบสอบถาม

งานวิจัยชิ้นนี้ตอบคำถามการวิจัยที่ว่า "ดัชนีวัดความยากง่ายแบบอิงคำศัพท์มีความเที่ยงตรงในระดับใด" ด้วยขั้นตอนในการวิจัยจำนวน 3 ขั้นตอน 1) การตรวจสอบผลกระทบของสัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อย คลังคำศัพท์ และเปอร์เซ็นต์ของความรู้คำศัพท์ในแต่ละบทอ่านที่มีแต่บทอ่าน 2) การสร้างดัชนี และ 3) การหาความเที่ยงตรงของดัชนี

ขั้นตอนที่ 1: การวิเคราะห์ความแปรปรวนแบบสามทางถูกนำมาใช้เพื่อตรวจสอบผลกระทบของสัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อย คลังคำศัพท์ และเปอร์เซ็นต์ของความรู้คำศัพท์ในแต่ละบทอ่านที่มีแต่บทอ่าน ผลของการวิเคราะห์แสดงให้เห็นว่าสัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อย ขนาดของคลังคำศัพท์ และเปอร์เซ็นต์ของความรู้คำศัพท์ในแต่ละบทอ่านที่มีแต่บทอ่านมีผลต่อความเข้าใจในการอ่าน นอกจากนี้ปฏิสัมพันธ์ระหว่างสัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อยและคลังคำศัพท์ก็มีผลกับความเข้าใจในการอ่าน อย่างไรก็ตามผลจากการวิเคราะห์ความแปรปรวนแบบสามทางไม่ได้แสดงให้เห็นว่าปฏิสัมพันธ์ระหว่างตัวแปรตัวอื่นๆ มีผลต่อความเข้าใจในการอ่าน ผลจากการวิเคราะห์นี้แสดงให้เห็นว่าตัวแปรที่ควรจะถูกใช้ในการสร้างดัชนีในขั้นตอนที่สอง คือ สัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อย ขนาดของคลังคำศัพท์และเปอร์เซ็นต์ของความรู้คำศัพท์ในแต่ละบทอ่าน และปฏิสัมพันธ์ระหว่างสัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อยและคลังคำศัพท์

ขั้นตอนที่ 2: เนื่องจากคำว่าความยากง่ายในงานชิ้นนี้หมายถึงระดับของความยากง่ายของบทอ่านสำหรับนักศึกษาในการทำความเข้าใจบทอ่านเหล่านั้น ดัชนีจึงถูกแบ่งออกเป็น 2 ส่วน 1) การบ่งชี้ความยากง่ายของบทอ่าน และ 2) การทำนายความเข้าใจในการอ่าน ความยากง่ายของบทอ่านด้วยถูกบ่งชี้โดยสัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อย โดยพิจารณาจากเปอร์เซ็นต์ของคำในแต่ละบทอ่านที่มาจากกลุ่มคำศัพท์ 3 กลุ่มแรกที่เกิดขึ้นบ่อย ส่วนความเข้าในการอ่านถูกทำนายโดยคะแนนจากข้อสอบคลังคำศัพท์และเปอร์เซ็นต์ของความรู้คำศัพท์ในแต่ละบทอ่านโดยใช้สมการถดถอย ผลที่ได้จากสมการจะอยู่ในรูปของคะแนนความเข้าใจในการอ่าน เมื่อคะแนนที่ได้ถูกแปลงเป็นระดับความยากง่ายของบทอ่านได้ 5 ระดับ ได้แก่ ยากเกินไป ยาก เหมาะสม ง่าย และง่ายเกินไป

ขั้นตอนที่ 3: ความเที่ยงตรงของดัชนีถูกตรวจสอบในรูปของความเที่ยงตรงเฉพาะหน้า ความเที่ยงตรงตามสภาพ และความเที่ยงตรงเชิงพยากรณ์ เพื่อที่จะตอบคำถามการวิจัยที่ว่าดัชนีวัดความยากง่ายบทอ่านแบบอิงคำศัพท์มีความเที่ยงตรงแค่ไหน แบบสอบถามแบบมาตราประมาณค่าแบบ 5 ระดับถูกใช้เพื่อวัดความเที่ยงตรงเฉพาะหน้า สหสัมพันธ์แบบสเปียร์แมนและสหสัมพันธ์ของเพียร์สันถูกนำมาใช้เพื่อหาความสัมพันธ์ระหว่างการบ่งชี้ความยากง่ายโดยสัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อย ผู้สอน นักศึกษาและสูตรวัดความยากง่ายจำนวน 2 สูตรเพื่อศึกษาความเที่ยงตรงตามสภาพของดัชนี ความเที่ยงตรงเชิงพยากรณ์ของการทำนายความเข้าใจในการอ่านโดยคลังคำศัพท์และเปอร์เซ็นต์ของความรู้คำศัพท์ในแต่ละบทอ่านที่มีแต่บทอ่านถูกตรวจสอบด้วยการใช้สหสัมพันธ์ของเพียร์สันเพื่อหาความสัมพันธ์และหว่างคะแนนและระดับความยากง่ายของบทอ่านที่ถูกทำนายกับคะแนนของนักศึกษาและการประเมินระดับความยากง่ายของบทอ่านของนักศึกษา

ผลจากการตรวจสองความเที่ยงตรงของดัชนีแสดงให้เห็นว่าดัชนีเป็นไปในทางที่ดี อย่างไรก็ตามดัชนีนี้มีปัญหาในเชิงปฏิบัติค่อนข้างมาก วิทยานิพนธ์นี้ได้นำเสนอข้อจำกัดของดัชนีและข้อเสนอแนะในการวิจัยในอนาคต

PATTEERA THIENPERMPOOL : A NEW VOCABULARY-BASED READABILITY INDEX FOR THAI UNIVERSITY STUDENTS. THESIS ADVISOR : ASSOC. PROF. JEREMY WILLIAM WARD, Ph.D., 351 PP.

READABILITY/LEXICAL FREQUENCY PROFILE/VOCABULARY SIZE/TEXT-SPECIFIC VOCABULARY KNOWLEDGE

This study aimed to devise a vocabulary-based readability index and to investigate the validity of the index. There were six groups of participants taking part in the study. Each of these groups consisted of 102, 45, 80, 5, 30 and 6 respectively. The data from the first group were used to devise the vocabulary-based readability index and the data from the latter groups were used to validate the index. There were four investigated variables: Lexical Frequency Profile (LFP), vocabulary size, text-specific vocabulary knowledge (TSVK) and reading comprehension. The instruments were 1) RANGE; 2) a yes/no test; 3) self-reports on unknown words; 4) translation tests; 5) reading comprehension tests and 6) questionnaires.

In order to answer the research question "To what extent would a purely vocabulary-based readability index be valid?", the present study was divided into three main stages: 1) an investigation of the effects of LFP, vocabulary size and TSVK on reading comprehension; 2) the development of the index and 3) validation of the index.

Stage 1: A three-way ANOVA was used to investigate the effects of LFP, vocabulary size and TSVK on reading comprehension. The results showed that there were some differences between the mean reading scores when LFP, vocabulary size

and TSVK were varied. Also, the interaction between LFP and vocabulary size had some effects on reading comprehension. However, the three-way ANOVA results revealed no effects of the interaction between LFP and TSVK, vocabulary size and TSVK, and LFP, vocabulary size and TSVK on reading comprehension. This showed that the variables should be used in Phase 2 should be LFP, vocabulary size, TSVK and the interaction between LFP and vocabulary size.

Stage 2: Based on the definition of readability, the level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts, the vocabulary-based readability index was comprised of two main parts: 1) indication of text difficulty and 2) prediction of reading comprehension. Text difficulty was indicated by LFP. Percentage of tokens from the first three frequency bands was used to indicate text difficulty. The reading comprehension was predicted by vocabulary size and TSVK by regression equations. The results from the equations were in terms of reading scores. These reading scores were converted into five bands of text difficulty: too difficult, difficult, optimal, easy and too easy.

Stage 3: The index was validated in terms of face validity, concurrent validity and predictive validity in order to answer the research question "to what extent would a purely vocabulary-based readability index be valid? A questionnaire with 5-point rating scales was used to investigate face validity. In order to investigate concurrent validity of LFP as an indicator of text difficulty, correlation coefficients between the indication of text difficulty by LFP, teachers, students and two traditional readability formulas were calculated by Spearman rho and Pearson *r*. The predictive validity of the prediction of reading comprehension at different vocabulary size and TSVK was investigated by

exploring the relationship between the predicted scores and bands of text difficulty and the students' actual scores and ratings of text difficulty by Pearson *r*.

The results from the validation of the index seem promising. However, the index tends to have a massive practicality problem. The constraints of the index and suggestions for further studies are presented in the dissertation.

School of English Student's signature ____________________

Academic Year 2009 Advisor's signature____________________

ACKNOWLEDGEMENTS

I am indebted to a number of people who have helped me in completing this thesis and would like to take this opportunity to thank them all.

Firstly, I would like to express my deep gratitude to Commission on Higher Education who has sponsored me with an MA and PhD scholarship and my colleagues in the Faculty of Education at Silpakorn University who gave me permission for study leave and moral support. Secondly, I would like to express my heartfelt gratitude to Phil Scholfield who gave me a chance to be an occasional student at University of Essex, United Kingdom. Without his help, I would not have had invaluable academic experience and a wonderful life in Colchester. In addition, I would like to thank him for his help, moral support and indispensable comments. Thirdly, I am grateful to my internal examiners, Assoc. Prof. Dr. Kanit Khaimuk, Dr. Sarit Srikhao and Dr. Peerasak Siriyothin, who spent their valuable time on reading and commenting on this thesis. I also appreciate all the help they have kindly given to me. Additionally, I would like to express my gratitude to Assoc. Prof. Dr. Richard Watson Todd for opening my world of research and for giving me encouragement, advice and guidance. Without his kind assistance, I would not have completed this thesis. I am also grateful to Paul Meara, Tom Cobb, and Wilfried Admiraal for their help with constructing and scoring the yes/no test. Next, I acknowledge my deep gratitude to my colleagues in Phil’s group discussion for the help and support. Also, I would like to express my cordial thanks to Thai postgraduates at University of Essex and the participants in Thailand who participated in the pilot and the actual studies for their co-operative performance although their names are not in this thesis because of the limited space. My thanks also go to my friends who have been constantly supported and encouraged me to persevere with my thesis. Last but not least, my thanks also go to my parents and my sister who are always generous with their moral support.

Patteera Thienpermpool

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

Page

TABLE OF CONTENTS (Cont.)

Page

TABLE OF CONTENTS (Cont.)

Page

TABLE OF CONTENTS (Cont.)

Page

TABLE OF CONTENTS (Cont.)

TABLE OF CONTENTS (Cont.)

TABLE OF CONTENTS (Cont.)

Page

TABLE OF CONTENTS (Cont.)

TABLE OF CONTENTS (Cont.)

Page

TABLE OF CONTENTS (Cont.)

Page

TABLE OF CONTENTS (Cont.)

TABLE OF CONTENTS (Cont.)

Page

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
BNC	British National Corpus
LFP	Lexical Frequency Profile
SD	Standard Deviation
TSVK	Text-Specific Vocabulary Knowledge
VLT	Vocabulary Levels Test

CHAPTER 1

INTRODUCTION

This chapter aims to present a general background of the study, which shows importance of reading and teachers' and students' ways of selecting texts. It also provides information on problems concerning traditional readability formulas which shows the need for another method of assessing the readability of texts. Additionally, the objectives, research questions and the limitations of the study are presented. It ends with the definitions of key terms used in the study.

1.1 Background of the Study

Reading seems to be one of the most realistic communicative skills in EFL context (Mikulecky and Jeffries, 1986) because most EFL students rarely need to speak the language in their everyday life but may need to read in order to access the information written in English (Eskey, 2005). It is generally defined as "a process whereby one looks at and understands what has been written" (Williams, 1984). It plays an important role in second or foreign language learning and teaching situations (Richards and Renandya, 2002). There are a number of reasons for this. Firstly, many students often have reading as one of their goals in learning the language because they want to be able to read for information, for pleasure, for their career, and for study purposes (ibid.). Secondly, it is stated that reading academic texts is considered one of the most important skills that ESL or EFL university students should master (Levine, Ferenz and Reves, 2000) because reading makes a contribution to other subjects. This is supported by research which shows that reading opens the door to learning about math, history, science, literature, geography and much more (U.S. Department of Education, www, 2004). Thus, students who are capable readers are likely to succeed in these subjects and develop confidence in their own abilities. Lastly, some students can gain good linguistic models of the language (Richards and Renandya, 2002) and learn some new vocabulary when they read (Nation, 2001). With these reasons, reading is integrated in language classes and as normally seen in commercial English textbooks such as Lifelines (Hutchinson, 1997), New Hotline (Hutchinson, 1998), New Interchange Students Book 3 (Richards, Hull, and Proctor, 1998) and Passages Student Book 1 (Richards and Sandy, 2008), reading skills are always included.

When reading is taught in ESL or EFL classrooms, like other skills, teachers need to plan what and how they are going to teach before teaching and students are encouraged to practice the skill outside class. One of the processes of teaching and learning reading that is worth discussing is the selection of texts.

For teachers, the selection of texts is a professional process (International Reading Association, www, 1994). The teachers need to select texts that will be used in a language class very rigorously and carefully. It is important for them to use a reliable and objective way to match appropriate texts to students. Texts used for reading instruction ought to be ones that students can read well but have not mastered (Lazar, 2004). If a text is too difficult for the students, the students are likely to focus too much on figuring out unknown words in the text (Taberski, 2000). They tend to struggle, become frustrated, give up trying to tackle the text, and say, “I just didn’t understand that” (O’Malley and Chamot, 1990). On the other hand, if a text is too easy, the students will be bored because they do not learn anything from the texts. Hence, the difficulty levels of instructional texts, which are selected, should be appropriately matched to students' capabilities (Carrell, 1987) in order to promote learning.

There are three main ways of selecting texts used by language teachers. The most popular way is that language teachers normally use their own experience and intuition to select texts for students or ask for feedback from learners (Klare, 1974-1975). If they find out that the texts are too difficult or too easy for the students, the teachers will select other texts for students in the next term. This is prominently used by most of the language teachers because it is an easiest way and it does not take much time. However, it is very subjective because the selection is completely based on teachers' judgment. This method is sometimes considered unreliable. The teachers tend to be continually faced with the problem of choosing materials appropriate to the needs of each learner or group (Harrison, 1977).

Another one is that language teachers can set text selection criteria or guidelines. This method has been used in English language teaching of some schools and institutions such as Department of Rhetoric & Writing, College of Liberal Arts, University of Texas at Austin (www, 2006), Victorian Association for the Teaching of English (www, 2008) and Maryland Institute College of Art (www, 2010). It is more reliable than the former one but it is still subjective because criteria are set by teachers or course designers. The teachers still need to use their own experience to select elements or factors which need to be included in the criteria. After deciding what to include in the criteria, the teachers need to write the criteria rigorously in order to make sure that the criteria are clear enough for other teachers to understand and use them to select a suitable text. This method takes a lot of time because one set of criteria may be suitable for only one particular subject or course. Additionally, readers who are one of the most factors in reading hardly take part in the design of the criteria so nobody can guarantee that the texts that are selected based on the criteria are so suitable that students can comprehend the texts.

The other way of text selection is done by indicating or estimating readability – level or degree of the ease or difficulty of texts for students to comprehend texts. According to Chall (1984), there are more than fifty readability formulas or indices such as Coleman-Liau Readability Formula, Dale-Chall Readability Formula, Flesch Reading Ease Readability Formula, Flesch-Kincaid Readability Formula, FORCAST readability formula, Fry Graph Formula, Gunning Fog Readability Formula, Kincaid formula, Powers, Sumner, Kearl Readability Formula (PSK formula), Rate Index (RIX) readability formula, SMOG readability formula; Spache Readability Formula, Homan-Hewitt Readability Formula, Linsear Write Readability Formula, Rayor Estimate Graph, and McAlpine EFLAW Test. Among these formulas, the most commonly used formulas are Flesch Reading Ease Readability Formula and Flesch-Kincaid Readability Formula. These two formulas are presented as the readability statistics in Microsoft Word. The formulas have been developed and published so as to reliably and objectively give a statistical analysis of the difficulty of texts. Unfortunately, they still have several drawbacks which are presented in the next section.

For students, text selection is also part of their learning process. When the students have learned reading in class, they are normally encouraged to practice reading outside class in order to promote self-study. They may select a text and read it for pleasure or for specific information, etc. in order to practice their reading. When they select a text, they usually select one based on their interests. Some of them may look for an interesting book or text in a library, a self-assess learning centre, or a bookstore while others surf the net in order to look for a text that interests them. Nevertheless, it is questionable whether the students can choose a text that they can comprehend. It is good if they can select a text that is suitable and comprehensible for them. However, some of them may select a text that is too difficult or too easy for them. If the text is too difficulty, the text will discourage the students (Higgins, 2009) and form bad attitudes towards reading. They may find out that there are too many new things to learn and decide to give up. On the other hand, if the students choose to read an easy or familiar text, they will be able to develop their reading rates or their reading fluency, learn some information from the text and increase their confidence in reading. Unfortunately, they will not learn much new vocabulary.

In conclusion, the term “text selection” in this study refers to the selection of texts in language learning and teaching employed by two groups of people: 1) teachers willing to select appropriate texts for students and 2) students looking for appropriate texts for themselves.

From the aforementioned methods of text selection, we have seen that the traditional readability formulas or indexes seem to be the only objective measure of readability at the moment. Unfortunately, the traditional formulas have some drawbacks. These drawbacks are presented in 1.2 so as to show that there should be another attempt to devise another readability formula or index that might not suffer from these drawbacks.

1.2 Statements of the Problems

It may be questionable why this study needs to be conducted because the aforementioned traditional readability formulas provide an easy and quick way to predict readability and as a result are a practical solution to the problem of estimating readability (Anderson and Davison, 1988 quoted in Meyer (2003); Klare, 1963; 1974-1975). In addition, it is claimed that the inclusion of other factors in the formula contributes more work than it improves the results and counting more things does not make the formula any more predictive of reading ease but takes a lot more effort (Klare, 1976). However, it is still questionable whether the traditional formulas provide a valid estimation of readability because of their four major drawbacks. The drawbacks are concerned with elements employed in the calculation; subject-specific factors; accuracy of the results; and applicability of the results. These drawbacks are discussed in detail below.

1.2.1 Elements Employed in the Calculation

In the past 60 or 70 years there have been several attempts to devise objective formulas to measure readability of texts. The author(s) of these formulas put a lot of effort to objectively measure readability. As a result, the formulas only measure variables that can be measured objectively. These variables are the surface structures of the text such as the number of words in the sentences and the number of letters or syllables per word. Unfortunately, they ignore readers who are the most important factor in reading. They estimate the level of difficulty of language use in a text rather than how difficult the text is for each reader. In other words, the traditional formulas roughly tell us whether the language use in a text is complicated or not but they cannot give any idea whether the level of difficulty of the text is suitable for a particular reader or not. By way of illustration, let us look at two sentences below.

> **Sentence 1:** Yesterday I went to an Elephant and Crocodile Farm in Nakhonpathom with my mother, my father and my younger sister by an air-conditioned bus because we wanted to see elephants and crocodiles.
>
> **Sentence 2:** Factory life necessitated a more regimented schedule, where work began at the sound of a bell and workers kept machines going at a constant pace (Sharpe, 2006).

The average numbers of words per sentence of these sentences are 34 and 27 respectively and the average numbers of letters per word are 4.9 and 4.7 respectively. From these average numbers, we can see that the first sentence consisting of 34 words is longer than the second one consisting of 27 words and the words in the first sentence tend to be longer. Hence, according to word length and sentence length, the first sentence is more difficult. Similarly, Flesch readability formula yields reading ease scores of the two sentences as 30.4 and 47.8 respective and Flesch-Kincaid readability test predicts that these sentences have the same level of difficulty which is 12^{th} American graders can read both sentences. However, the first sentence was written by a Matthayom 1 student while the second sentence was taken from a TOEFL practice book. This shows that the estimations of word difficulty by word length and sentence complexity by sentence length do not properly reflect the actual difficulty of sentences or texts.

Furthermore, there is no guarantee that these surface elements employed by the traditional readability formulas really affect readability of texts. Davison and Kantor (1982), aiming to investigate whether the readability formulas actually define readability, compare two versions of four texts: the original versions intended for adult readers and the adapted versions intended for less skilled readers. Some changes such as splitting complex sentences into component clauses, changing vocabulary items are made in order to make the texts easier to read. These changes are made to make the text conform to a certain level of readability estimated by the traditional formulas. However, it was found out that the changes are not always the most successful and some actually make the texts harder to understand. This shows that the measured elements of the formulas like word length and sentence length are not the actual features of texts that make them easy or difficult to understand.

1.2.2 Applicability of the Results

The results of most of the readability formulas such as Automated Readability Index, Dale-Chall readability formula, Flesch-Kincaid readability formula, FORCAST readability formula are in terms of American grade level. It may be difficult to apply the results to other language learning and teaching situations especially second or foreign language learning and teaching situations because English is used as the first language in America. Students have a lot of chance to expose to authentic English while students in some countries rarely experience any English in their everyday life. In addition, the language learning and teaching situations outside America are different from the ones in America. There is no research stating that American grade levels are equivalent to grade levels in any country including United Kingdom where English is also used as the first language (Allan, McGhee and Krieken, www, 2005). If a formula results in American grade level of 9, then 9th grade students in America would only just be able to understand that piece of text. It does not mean that students all over the world who are studying in grade 9 will be able to understand this text. Furthermore, there is sometimes a further question of grade level that where does a piece of writing fit on a reading grade or reading age continuum? (Klare, 1974-1975). Therefore, the way of presenting the results of the readability needs to be improved in order to make the results applicable to other language learning and teaching situations.

1.2.3 Subject Specific Factors

Many subjects have terms, which have particular meanings within the field of knowledge. Students in a particular field are required to learn some of the subject-specific terms in that field. For example, the word "dermatitis", which means a disease in which the skin is red and painful (Cambridge Dictionaries Online, www, 2010) is a technical term in the field of medicine. When we look at the available readability formulas such as Fog index and SMOG readability formula, we can see that such words would have a significant impact on the overall score of the formulas, which counted the total number of syllables and/or number of words with three or more syllables, for a particular text because this word is treated as a long and difficult word. However, it should not be difficult for most students in medicine. Hence, subject-specific terms may distort the estimation of some of the readability formulas and indices which are based on the total number of syllables and/or number of words with three or more syllables.

1.2.4 Consistency of the Results

Although it is very objective for the traditional readability formulas to measure surface structures of the text and most of the prominent readability formulas are computerised, there has been some criticism of consistencies of word counts and syllable counts. For example, Flesch readability formula, Flesch-Kincaid readability test and Gunning Fog readability formula rely heavily on word length and/or sentence length which require researchers to count number of words in the text and word length which requires researchers to count number of syllables. Both sentence length and word length can be objectively measured and the counts are usually done by computer but their results are, unfortunately, not consistent. The researcher tested readability of three texts by using Microsoft Word 2003 and uploading them to two online readability tests which are http://www.onlineutility.org/english/readability_test_and_improve.jsp and http://juicystudio.com/ services/readability.php. Microsoft Word 2003 yields the results of Flesch Reading Ease and Flesch-Kincaid Grade Level. The results from these two formulas from Microsoft Word 2003 are surprisingly different from the results from the online formulas. A webmaster of online readability calculators (http://juicystudio.com/services/readability.php), Adamovic, (personal communication, May 7, 2007) clarifies that he recognised the differences of the results and reveals the reason for the difference that the ways of splitting words into syllables are different. Hence, it is questionable which program employs the most suitable word and syllable counts and which one yields the most accurate results.

From these four drawbacks (see 1.2.1, 1.2.2, 1.2.3 and 1.2.4), we can see that there is a need to devise another readability formula or index. In the hope that these drawbacks would be resolved, the index was devised by employing both text-based variable (Lexical Frequency Profile or LFP) and reader-based variables (vocabulary size and text-specific vocabulary knowledge or TSVK) in order that the index would measure how difficult a text is for each reader more accurately and it would present results that are more applicable than the available readability formulas or indexes. By using students' or readers' vocabulary .knowledge in a text called text-specific vocabulary knowledge (TSVK), it should be possible to account for subject specific knowledge. By using variables that can be measured objectively without any subjective guidelines or criteria, it may be possible that results from the devised index would be more consistent than the available formulas.

1.3 Purpose of the Study and Research Question

This study aims to devise a readability index called vocabulary-based readability index and investigate its validity. The devised index should be possible to help language teachers and Thai university students predict level or degree of the ease or difficulty of a text for individual students to comprehend the texts in order that they can select texts that they can comprehend. This study therefore aims to find answers to the following question:

- To what extent would a purely vocabulary-based readability index be valid?

From the name of the index, we can see that the development of the index is based on vocabulary. There are three main elements of vocabulary which are 1) lexical frequency profile (LFP) or proportion of low and high frequency words in a text; 2) vocabulary size or students' or readers' general vocabulary knowledge and 3) text-specific vocabulary knowledge (TSVK) or students' knowledge of vocabulary in a text. These elements can be categorized into two types of variables: 1) a text-based variable which is LFP and 2) reader-based variables which are vocabulary size and TSVK.

Correspondingly, when we combine these elements to the definition of readability defined in 2.3 as the level or degree of the ease or difficulty of texts for

Thai university students to comprehend the texts, the index will be comprised of three components: 1) the use of LFP, which is the text-based variable, to indicate the level of text difficulty and 2) the use of vocabulary size to predict reading comprehension scores and how easy or difficult a text is for individual students to comprehend the text and 3) the use of TSVK to predict reading comprehension scores and how easy or difficult a text is for individual students to comprehend the text.

In order to answer the research question, face validity of the three components of the vocabulary-based readability index, concurrent validity of the use of LFP to indicate text difficulty and predictive validity of the prediction of reading comprehension at different vocabulary size and at different TSVK were investigated.

1.4 Limitations of the Study

- Most of the participants taking part in the present study majored in science-oriented fields so it is unknown whether the results were applicable or generalizable to students majoring in non-science-oriented fields.

- The participants taking part in the study were students in public universities so the findings may not be generalizable to students in private universities.

- Several variables, namely text type, text length, sentence length, that may affect reading comprehension text difficulty were controlled. It might be problematic if we apply or generalize the vocabulary-based readability index to texts with other text types, text lengths and sentence lengths.

- The term “reading comprehension” is defined in Chapter 2 as the level or degree of readers’ understanding of a text in terms of scanning and paraphrasing and the reading comprehension test aims to test only the ability to scan and paraphrase so the applicability of the vocabulary-based readability index to other levels of reading comprehension may be questionable.

1.5 Definitions of Terms

The following terms, which are used throughout the present investigation, are presented together with the working definitions.

Lexical Frequency Profile (LFP) refers to proportion of low and high frequency words in a text.

Readability refers to the level or degree of the ease or difficulty of texts for readers to comprehend the texts.

Reading comprehension refers to the level or degree of understanding of a text in terms of scanning and paraphrasing**.**

Text-specific vocabulary knowledge refers to reader’s knowledge on vocabulary in a text.

A **token** or running word is any occurrence of a word form in the text, regardless of whether it is occurring for the 1st or the nth time.

A **type** is any word form which is counted only once, regardless of how many more times it might occur.

Vocabulary is defined as all the words which exist in English

Vocabulary size is quantity of reader's general vocabulary knowledge in terms of recognition of written form of words and meanings.

1.6 Summary

In this chapter, a description of the background of the study has been given in order to provide an overview of roles of reading and traditional ways of text selection. This was followed by statements of the problems stating why a new index should be devised. The objectives, the research question and the limitation of the study are then presented. Lastly, the working definitions of terms used in the study are given.

CHAPTER 2

REVIEW OF RELATED LITERATURE

This chapter presents a review of the literature related to the present study. According to Chapter 1, the purpose of the study is to devise a readability index called vocabulary-based readability index and investigate its validity. This chapter then presents nature of reading, reading comprehension, readability and the literature relating to how the vocabulary-based readability index was devised and the design of the index including importance of vocabulary, the relationship between reading and vocabulary.

2.1 Nature of Reading

Although a definition of reading tends to be given at the beginning of texts about reading or the teaching of reading, the act of reading is not completely understood nor easily described (Aebersold and Field, 1997) in the definition because "different people use the term "reading" in different ways" (Nuttall, 2005). By way of illustration, let us look at four situations from Wallace (1992).

1. An adult is having a sight test at an optician's and is asked to read a list of words.
2. A child in a class is shown a flash card with the word 'here' on it by her teacher.
3. An Islamic religious leader asks a congregation of boys to read aloud the Koran.
4. The owner of a new computer asks an experienced friend about the instructions in the manual.

In the first situation, when the adult is asked to read a list of words, he/she is asked to identify the words in the list by reading aloud in order to examine whether the adult has the ability to see the words. This can be done with or without an understanding of the words (Richards and Schmidt, 2002). In the second situation, the teacher is likely to ask the child to decode the text rather than attribute a meaning to the word. In other words, the child is asked to recall and recognize the individual phonemes and phoneme blends that are represented by the printed words (Improve-Reading-Skill.com, www, 2005-2007). In the third situation, the text, which is the Koran, is learned and recited by rote. The boys can recognize sections of the text according to some features of the text such as the position of print on the page and the headings. They may not be able to render the same section aloud if they encounter in a different textual context. This reading process is called recitation. In the fourth situation, the experienced friend needs to perceive the manual or the written text in order to understand the meaning of the text (Richards and Schmidt, 2002).

These four reading situations can be categorized into three groups of possible words used in definition of reading in Nuttall (1982). These categories are: 1) articulate, speak, pronounce, etc. 2) decode, decipher, identify, etc. and 3) understand, interpret, meaning, sense, etc. We can see that the first situation falls into the first category. The second and the third situations seem to be part of the second category and the fourth one tends to fall into the third category. Although the words used to describe these four situations are different, it can be concluded that the term “reading” is something that involves the reader, the text, and the interaction between the reader and the text (Rumelhart, 1977).

2.2 Reading Comprehension

Reading comprehension is one of the five key components of essential reading instruction as identified in the National Reading Panel Report (National Institute of Child Health and Human Development, 2000). Surprisingly most of the studies on reading comprehension such as Perkins (1984); Stavans and Oded (1993); Qian (1999); Geiger and Millis (2004); Razi (2005) hardly define what reading comprehension is because it is assumed that we all know what it is (Wilhelm, www, 1996-2010). However, the term "reading comprehension" has been defined in several ways as follows:

- Reading comprehension is defines as a part of communication process of getting the thoughts that were in the author's mind into the reader's mind (Fry, 1965).

- Reading comprehension refers to whatever their assessment instrument measured, for example, number of ideas recalled in a free recall task, number of correct answers on a multiple choice recognition test of memory for text contents (Carroll, 1972).

- Reading comprehension can be defined in terms of the coherence of the representation the reader constructs and – depending upon the reader's goal – the relation between the reader's representation and the representation intended by the author (Lorch Jr., van den Broek, 1997).

- Comprehension is described as the ability to find meaning in what is read (Mohamad, 1999).

- Reading comprehension involves constructing meaning that is reasonable and accurate by connecting what has been read to what the reader already knows and thinking about all of this information until it is understood (Learning Point Associates, www, 2004).

- Reading comprehension is the process of constructing meaning from text (Lenz, www, 2005)

- Reading comprehension is defined as the degree to which we understand what we read (ReadingIsGood, Ltd., www, 2008).

- Reading comprehension refers to understanding and creating meaning from written materials (How To Do Things, www, 2009).
- Reading comprehension refers to constructing the meaning of the oral or written messages (Iwai, www, 2010).
- Reading comprehension refers to the level of understanding of a passage or text (Pakhare, www, 2010).
- Reading comprehension is defined as the level of understanding of writing (Wikipedia, www, 2010).
- Reading comprehension can be defined as the ability to understand information presented in written form (Center for Advancement and Learning, Muskingum College, www, n.d.).

From the aforementioned definitions of reading comprehension, some words, such as reader, we, author, read, understanding, constructing meaning and text are frequently mentioned. These words can be categorized into four groups: 1) reader; 2) author; 3) action of reading comprehension (read, understanding and constructing meaning); and 4) texts. This can be concluded that reading comprehension involves readers understand texts written by author. Furthermore, a couple of them refer to reading comprehension as a process of getting the thoughts and constructing meaning from texts.

In order to select the most suitable definition for the study, it is important to look at how the reading comprehension involves the present study. Based on the purpose which is to devise a readability index and investigate its validity (see 1.3), the key term used in the present study is "readability", being defined as the level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts (see 2.3). According to British National Corpus (BNC), the word "comprehend" in the definition is in the same word family as the word "comprehension". That is to say, it is the headword with seven family members, namely, comprehends, comprehended, comprehending, non-comprehending, uncomprehending, uncomprehendingly, comprehension and comprehensions. The word “comprehension” is one of its family members. The word “comprehend” is a verb, showing an action done by readers to the texts. According to Cambridge Dictionary Online (www, 2010), the word “comprehend” is defined as to understand something completely. This shows that it seems to have something to do with understanding of texts or constructing of texts rather than a process. Therefore, this term “reading comprehension” is defined in the present study as the level or degree of readers’ understanding of a text.

When we talk about reading comprehension as the level or degree of readers’ understanding of a text, it is crucial to be aware that there are several types or levels of comprehension. Herber (1978) divides reading comprehension into three levels of comprehension questions based on Bloom’s levels of cognitive domain (Bloom, Engelhart, Furst, Hill and Krathwohl, 1956). These three levels are 1) literal questions requiring students or readers to recall or reorganize information explicitly presented in the reading material; 2) interpretive questions asking for an explanation, inference, conclusion, or summary; and 3) applied questions utilising the students’ or readers’ background knowledge and lead them to evaluate, elaborate, predict, or solve problems based on implicit information in the text.

Similarly, Mohamad (1999) also states that there are three levels of comprehension. However, there are some differences between levels of comprehension divided by Herber and Mohamad as follows:

1) Mohamad divides reading comprehension in terms of levels of comprehension rather than comprehension questions.

2) Mohamad calls the first two levels which are literal and interpretative comprehension in the same way as Herber (1978) but he calls the third level critical reading. Mohamad’s critical reading is very similar to Herber’s applied questions. It involves the understanding of text and the evaluation of ideas and information in the text. Students or readers who are asked to do critical reading are required to differentiate between facts and opinions, recognize persuasive statements and judge the accuracy of the information given in the text, etc.

3) Although Herber and Mohamad have similar ways of naming the levels of reading comprehension, the descriptions of the three levels are different. Based on Mohamad (1999), literal comprehension is simpler than Herber’s literal comprehension questions. It involves surface meanings only. When students read a text for literal comprehension, they are required to find information or ideas that are explicitly stated in the text while Herber’s literal comprehension questions also cover reorganizing information which is categorized as interpretative comprehension by Mohamad. Additionally, according to Mohamad, interpretative comprehension involves the ability to analyze what is read carefully, to see the relationships among ideas, to draw conclusions, to make generalizations and to predict outcomes while according to Herber, the ability to predict has something to do with applied questions.

Recently Day and Park (2005) also divide reading comprehension into types but they do it in more details than Herber (1978) and Mohamad (1999). Day and Park designed taxonomies of the types of comprehension as a checklist for language teachers and material developers. There are six types of comprehension in the checklist: 1) literal comprehension – an understanding of the straightforward meaning of the text; 2) reorganization requiring students or readers to use information from various parts of the text and combine them for additional understanding; 3) inference involving students or readers combining their literal understanding of the text with their own knowledge and intuitions; 4) prediction involving students or readers using both their understanding of the passage and their own knowledge of the topic and related matters in a systematic fashion to determine what might happen next or after a story ends; 5) evaluation requiring students or readers to give a global or comprehensive judgment about some aspect of the text; and 6) personal response requiring students or readers to respond with their feelings for the text and the subject.

In order to select the level or type of reading comprehension suitable for the study, it is important to consider the participants of the present study. Most of the participants taking part in the study were from Suranaree University of Technology. They studied English as a foreign language. They hardly read English texts in their everyday life and their reading proficiency ranged from beginners to pre-intermediate (see 3.1). Moreover, questions of interpretative and applied comprehension, reorganization, inference, prediction and evaluation tend to be more difficult than literal comprehension questions because they do not only require the test-takers or participants to understand the reading passages but they also involves combining their understanding with their own knowledge and intuitions (Day and Park, 2005). Therefore, literal comprehension seems to be the most suitable level for the participants.

According to Mohamad (1999) and Day and Park (2005), literal comprehension refers to understanding of surface or straightforward meanings of the text. This might involve skimming, scanning, paraphrasing, etc. However, in this study, the understandings of surface meanings refer to only scanning and paraphrasing for two reasons: 1) the students' level of reading proficiency was not high and 2) the textbook used for learning and teaching reading in their classes was World Class Readings 2 (Rogers, 2005). Most of the reading comprehension questions in this book require scanning and paraphrasing so the students seem to be familiar with scanning and paraphrasing. Hence, in the present study, the level of comprehension or readers' understanding of a text refers to the literal comprehension focusing on the ability to scan or look for specific information and paraphrase it only.

2.3 Readability

Readability has been studied since 900 A.D. (Abram, 1981 quoted in Taylor and Wahlstorm, www, 1999). However, it "continues to be among the most discussed, misunderstood, and misused concepts in reading" (Pikulski, www, 2002). There are several attempts to define the term "readability" as follows:

- Readability is generally defined as the study of matching reader and text (Gilliland, 1972).
- Readability refers to the ease of understanding or comprehension due to the style of writing (Klare, 1963; 1976).
- Readability refers to "the combination of structural and lexical difficulty" (Nuttall, 1982; 2005).
- Readability can be defined as features of text that cause difficulty to readers (Alderson and Urquhart, 1984).
- Readability refers to factors that make a text difficult (Wallace, 1992).
- Readability is "the ease with which a text can be read" (Aaron and Joshi, 1992).

- Readability is "a measure of the predicted difficulty of a text through reference to readability formulae of different kinds" (Davies, 1995).

- Readability is defined as the various aspects of a text that are likely to make it easy or difficult for a reader to understand and enjoy (Read, 2000).

- Readability can be defined as "the level of ease or difficulty with which text material can be understood by a particular reader who is reading that text for a specific purpose" (Pikulski, www, 2002).

- Readability refers to "measure of the ease with which a text can be read; usually expressed as a grade level" (University of Prince Edward Island, www, 2006).

- Readability refers to "the measure of how comfortably or easily your text can be read" (Gregory, www, n.d.).

- The term readability refers to "all the factors that affect success in reading and understanding a text". (Johnson, www, n.d.)

From a preliminary glance through the aforementioned definitions, we can see that the words which frequently occur in the definitions are "text", "ease", "difficulty", "measure", "reader", "understanding", and "comprehension". This implies that readability has something to do with the measurement of the ease or difficulty of texts that affects reader's understanding or comprehension. When we look through the definitions in more detail, it is found out that there are three main ways of defining readability. Firstly, it is defined as the level of ease or difficulty a text can be read. This idea is presented by Klare (1976); Aaron and Joshi, (1992); Pikulski (www, 2002). Secondly, some scholars such as Nuttall (1982); Nuttall (2005); Alderson and Urquhart (1984); Wallace (1992); Johnson (www, n.d.); state that readability is features or aspects that make a text easy or difficult. Lastly, this term is also defined by Davies (1995); Read (2000); University of Prince Edward Island (www, 2006); Gregory (www, n.d.) as a measure or study of the ease or difficulty with which a text can be read.

The above definitions are varied depending on the purposes of the studies. Since the devised index intended to help language teachers and Thai university students select texts which are suitable for students' reading comprehension, the term "readability" will be used in this study when referring to the level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts.

2.4 Available Readability Measures

The available methods for measuring readability may conveniently be arranged into five subsections which are 1) subject assessment; 2) objective question and answer techniques; 3) tables and charts; 4) cloze test; and 5) readability formulas or indexes (Gilliland, 1972). The methods that are ubiquitously used are cloze test and readability formulas. These two are described and discussed below.

2.4.1 Cloze Test

The cloze test is widely used in language assessment, particularly for the assessment of reading skills and overall scores (Bailey, 1998). A cloze test involves taking one or more reading passages, deleting every n^{th} word (a fixed ratio) and leaving a blank in its place. It is used to determine whether a participant or reader has difficulties with reading the passage(s). Test-takers or participants are required to read the passage(s), predict missing words from the context and write a suitable word in each blank (Read, 2000; Buck, 2001).

The cloze test was first applied to measure readability of texts by Taylor (1953 quoted in Read, 2000). It is based on the theory that the higher the participant's reading ability the greater the success of predicting the missing words. Therefore, if we know the reading ability of a representative sample of a group of population from a cloze test, the results of the test can then be used to determine the readability of the text (Allan et al., www, 2005).

In addition, it tells us whether a particular group of readers can comprehend the text well enough to complete the cloze test (Klare, 1976). If the text is well written, the readers should be able to fill in the blank based on the rest of the sentence. If the text is not well written, the readers will not be able to fill in the blank because the readers can't understand the sentence. However, the cloze test is criticised that some readers or test takers can fill in the gaps correctly because they are familiar with the patterns of the language. They do not need any ability to comprehend the text in order to get the correct answers.

2.4.2 Readability Formulas or Indexes

A readability formula or index "is like a yardstick that helps us measure certain qualities in the writing so we can make objective judgments about reading level" (Laubach and Koschnick, 1977 quoted in Taylor and Wahlstrom, www, 1999). It is concerned with semantic and syntactic aspects of the text (Allan et al., www, 2005). According to Chall (1984), there are more than fifty readability formulas or indexes which have been developed and published so as to give a statistical analysis of difficulty of texts. Readability was originally calculated by hand. At present, the readability formulas can be categorised into two main groups which are computerized and non-computerized readability formulas. Computerised formulas are, for example, Flesch readability formula; Flesch-Kincaid readability test and Passive sentences readability score. Non-computerised formulas are, for instance, Homan-Hewitt readability formula; Linsear Write readability formula; and Rayor estimate graph.

There are several attempts to computerise the readability formulas although there is only one explicit difference between computerised and non-computerised formulas which is the users' convenience of accessing to the formulas or indexes. In fact, both of them employ only word length and sentence length which are surface elements of reading texts that can be counted mechanically in calculation (see 1.2.1). They ignore readers who are one of the most important factors in reading. This study therefore aims to devise a vocabulary-based readability index which takes into account both the text-based and the reader-based variables.

Several of the aforementioned traditional readability formulas have been integrated into widely-used word processors such as Microsoft Word and Corel WordPerfect (Ownby, 2005). Microsoft Word generates readability statistics using three different methods which are passive sentences, Flesch readability formula and Flesch-Kincaid readability test. To determine readability of any document, we can use the grammar tool to check grammar. After the grammar check is completed, the readability statistics of the document will be displayed. For Corel WordPerfect, we can select Grammatik from the Tools menu and select the Options button. Then select analysis and readability. WordPerfect generates statistics on the Flesch-Kincaid readability test, passive sentences, sentence complexity and vocabulary complexity (ibid.). It also compares the scores of the document we are analyzing with another document available on WordPerfect (Vaso, www, 2005).

2.5 The Vocabulary-based Readability Index

From the name of the devised readability index, we can see that this index was devised based on vocabulary. Richards and Schmidt (2002) define vocabulary as a set of lexemes which includes single words, compound words and idioms because vocabulary consists of more than just single words (Read, 2000). However, from section 1.3, vocabulary in the present study has three main elements: LFP, vocabulary size and TSVK. All of these elements have something to do with single words rather than compound words and idioms so the term “vocabulary” in this study means all the words which exist in English.

There are two main types of vocabulary covering all aspects of what is involved in knowing a word (Nation, 2001). They are receptive and productive vocabulary. “Receptive vocabulary use involves perceiving the form of a word while listening or reading and retrieving its meaning. Productive vocabulary use involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form” (Nation, 2001). Since the present study aims to devise a readability index and it mainly involves reading comprehension, receptive vocabulary was employed.

It has been accepted that receptive vocabulary plays an important role in reading comprehension (Davis, 1968; 1972; Yap, 1979; Anderson and Freebody, 1981; Nagy, 1988; Brisbois, 1995; Alderson, 2000; Read, 2000; Boyle and Kirk, 2006; Golkar and Yamini, 2007; Cooper, 2008) in the first language (Sternberg 1987; West and Stanovich 1991), the second language (Horst, Cobb and Meara, 1998) and the foreign language because we have to be able to decode the printed message in order to understand text meaning (Adams, 2004; Alderson, 2000; Day and Bamford, 1998). Otherwise, we cannot understand any text without knowing what most of the vocabulary mean (Nagy, 1988) either in one’s native language or in a foreign language (Laufer, 1997).

It is believed that a reader's vocabulary relates to the person's reading comprehension (Anderson and Freebody, 1981; Nagy and Anderson, 1984 and Nagy and Herman, 1987). This is supported by numerous studies, such as Farley and Elmore (1992); Joshi and Aaron (2000); Qian (2002), showing the strength of the relationship between vocabulary and reading comprehension. On the one hand, having larger vocabulary sizes would assure better comprehension of texts (Horst, Cobb and Meara, 1998). On the other hand, poor readers have a smaller vocabulary size (Daneman, 1991) and their small vocabulary size impedes their reading comprehension (Pinnell, Lyons, Deford, Bryk, Seltze, 1994).

Although several studies reveal that readers' vocabulary knowledge strongly relates to their reading comprehension and readers' vocabulary knowledge is considered as the single predictor of how well readers can understand the text (Anderson and Freebody, 1981), none of the traditional readability formulas make use of it because they tend to ignore readers who read the text (see 1.2.1). The strong relationship between readers' vocabulary knowledge and reading comprehension leads to the possibility that vocabulary knowledge can predict how easy or difficult a text is for individual students to comprehend it. Therefore, vocabulary knowledge will be used in the development of the vocabulary-based index in order to take readers, who are one of the most important factors in reading, into account (see 2.5.2).

Most traditional readability formulas make use of the strong relationship between vocabulary and reading comprehension by measuring vocabulary difficulty because it is believed that the proportion of difficult words in a text seems to be the most powerful predictor of text difficulty (Anderson and Freebody, 1981). They determine the difficulty of text on the basis of vocabulary difficulty (Davey, 1988).

The vocabulary difficulty has been measured in the traditional readability formulas by three main methods which are 1) counting number of syllables or letters in Flesch readability formulas (Flesch, 1943; 1948); Farr-Jenkins-Paterson formula (Farr, Jenkins and Paterson, 1951); Gunning Fog's index or Fog index (Gunning, 1952); Fry's readability graph (Fry, 1965); SMOG grading (McLaughlin, 1969); 2) counting number of words that are in or not in well-known word lists like Dale and Chall (1948) and Thorndike-Lorge count (Thorndike and Lorge, 1944) in Lorge formulas (Lorge, 1939; 1948); Dale-Chall formula (Dale and Chall, 1948); Spache formula (Spache, 1953); Botel formula (Botel, 1962); Rogers formula (Rogers, 1962); and 3) determining mean noun frequency in Elley formula (Elley, 1969). However, although Schwartz, Sparkman, and Deese (1970) and Wang (1970 quoted in Klare, 1974-1975) states that readers can provide more sensitive and reliable judgments of the readability than the formulas using counts of language variables, unfortunately, most of the readability formulas do not take any readers' variables in the estimation of text difficulty (Ekwall and Henry, 1968; McLeod, 1962 quoted in Klare, 1974-1975).

To fill in this gap of the related literature on readability and to get rid of the drawbacks of the traditional readability formulas stated in 1.2, the development of vocabulary-based readability index relies on the use of vocabulary difficulty to predict text difficulty and the use of readers' or students' vocabulary knowledge, namely, vocabulary size and TSVK to predict reading comprehension including reading comprehension scores and how easy or difficult a text is for individual students to comprehend the text. The index will then be comprised of three components: 1) prediction or indication of text difficulty and 2) prediction of reading comprehension at different vocabulary size and 3) prediction of reading comprehension at different

TSVK. The development of the vocabulary-based readability index is illustrated in terms of theoretical framework below.

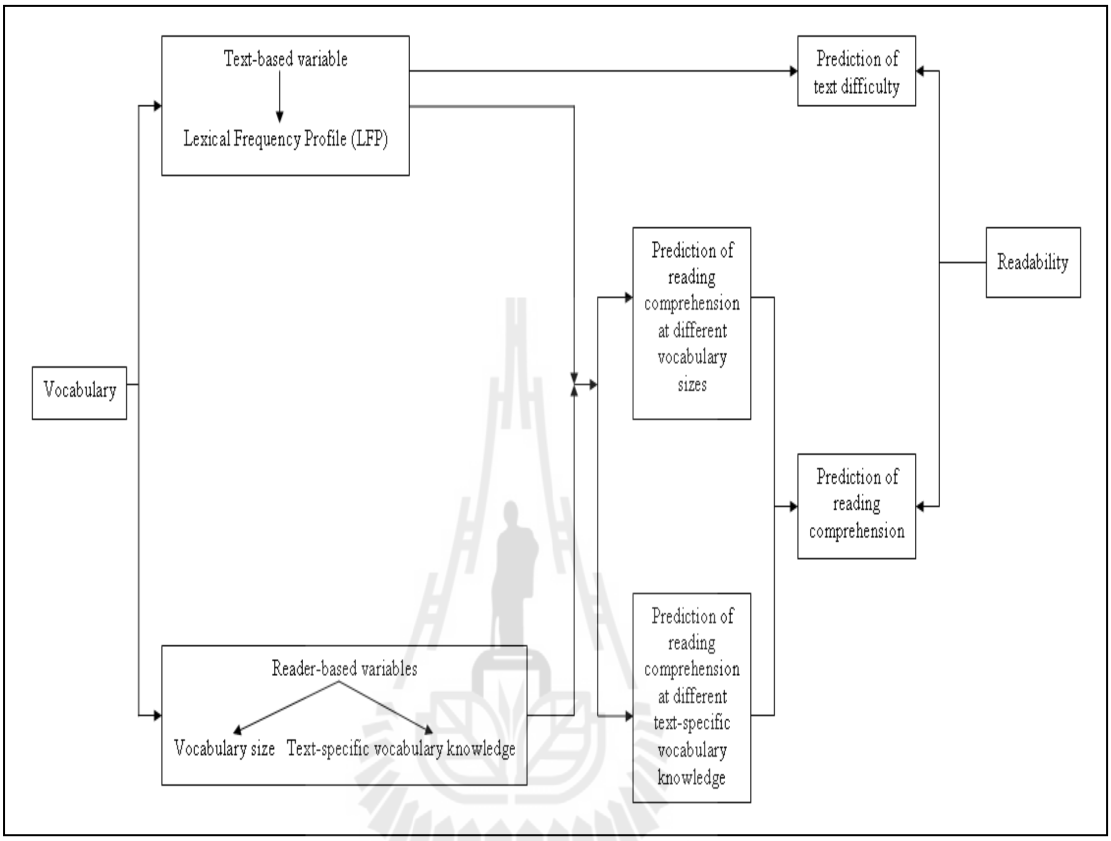


Figure 2.1: Theoretical Framework for the Development of Vocabulary-based Readability Index

From Figure 2.1, we can see that there are two types of variables relating to vocabulary. These variables are text-based and reader-based variables. The only text-based variable used in the study is LFP or proportion of low and high frequency words in a text. The reader-based variables are vocabulary size and TSVK. The text-based variable or LFP is used to predict text difficulty which is one aspect of readability. It is also used with the reader-based variables to predict the other aspect of readability including prediction of reading comprehension at different vocabulary size and prediction of reading comprehension at different TSVK. The literature underlying the above theoretical framework is presented below.

2.5.1 Prediction of Text Difficulty

Section 2.5 presents how important receptive vocabulary is in reading comprehension. However, text difficulty is not a matter of vocabulary only. There are many other factors influencing reading comprehension. These factors are the code or the written form in which the text has been expressed, interest and motivation, background knowledge, legibility of print, letters, illustration and colour, conceptual difficulty and syntax. These factors can be categorised into two main categories: reader factors and text factors. Reader factors involve interest and motivation, the code in which the text has been expressed and background knowledge. Text factors are legibility of print, illustration and colour, conceptual difficulty and syntax. These two categories will be presented in detail below in order to show why they were discarded when the vocabulary-based readability index was devised in this present investigation and why vocabulary is the main focus of the study.

The first reader factor which might affect text difficulty is interest and motivation. When we are motivated or interested in something such as texts or books, it is likely that we want to know about it or take part in it (Macmillan Dictionary, www, 2009). According to Gilliland (1972), there are two main issues when we talk about a text being motivating or interesting. One is about the enjoyment which arouses pleasurable feelings. This makes readers who are interested in the text have tolerance of difficulties while reading the text. In other words, they will try to do their best in order to read what they are interested in. For example, a Thai girl looks up many Korean words in a Korean-Thai dictionary because she wants to understand a

Korean text in order to find out what her favourite singer wrote in his/her online diary. The other issue concerning why readers may find a text interesting is when the text provides intellectual stimulus or assistance in solving problems. For example, a Thai lady whose English language proficiency is not very high reads a very long text about bird flu because she would like to know how to diagnose the disease, how humans can be infected and how to protect herself from the disease. Although interest and motivation encourage readers to concentrate on reading, this kind of information does not permit teachers or researchers to make an accurate prediction about text difficulty in respect of a particular individual (Gilliland, 1972). Therefore, this factor is not included in the study.

Another reader factor is the code or the written form in which it has been expressed. Readers find a text difficult because they are not familiar with the code in which it has been expressed. For example, readers who do not know any Finnish do not understand any texts written in Finnish because the readers are not familiar or do not know the written form of the language (Nuttall, 1982; 2000; 2005). This factor is not investigated in this study because it is so obvious that readers who do not know the written form of texts will find the texts difficult and incomprehensible.

The other reader factor is background knowledge. Background knowledge seems to be the most helpful with a moderately unfamiliar text (Roller, 1990). When readers are not familiar with a text, the ones who have some background knowledge about what they are reading tend to be able to predict and understand the texts more efficiently and faster than the ones who do not have any background knowledge. Although this factor has a big impact on reader's reading comprehension, it seems to be impossible to emphasize individual's background knowledge in order to predict text difficulty so this factor is not included in the study.

One of the text factors is legibility of print. Legibility refers to the aspects of typography which determine how readily the letters and words of the text will be deciphered (Gilliland, 1972). These aspects include such features as size of type and the particular type font, layout variables such as line length, size of margins, etc. For example, according to Watts and Nisbet (1974) and Tinker (1963), a point size in the range 14 point to 18 point is most effective. When a text contains less legible features and layout, the reader's motivation on reading the text may be lowered. Unfortunately, more legible texts will not turn a poor reader into a good one over night (Harrison, 1980). This shows that it seems unrealistic to expect that texts are easy to comprehend as a result of using more legible texts. Hence, it seems unlikely to use legibility of print to predict text difficulty.

Another factor relates to illustration and color (Gilliland, 1972). Most of the children's books contain colorful illustrations in order to motivate young readers. However, the usefulness of illustrations as an aid to learning is still questionable. Students are usually suggested to look at illustrations in order to skim a text and predict what the text is about before reading. Illustrations are very useful when they contain a great deal of information while some of them may not relate to the task. These illustrations may distract their attention or confuse them while reading. In addition, some pictures may not be clear. Some readers need to be able to understand the text in order to have an idea of what the ambiguous drawing means. For colors, it is also difficult to evaluate how colors affect text difficulty because colors might increase reader's motivation while it could distract reader's attention or lower legibility. We can see that illustrations and colors are matters of personal preference. Although it may somehow affect text difficulty, it is questionable how it does.

For conceptual difficulty, sometimes the vocabulary used in a text does not seem difficult and readers find it familiar because it occurs frequently in everyday life but most readers may find it difficult to understand because the concept to which it refers can only be fully understood by a particular group of people. For instance, the phrase “a black hole in space” (Harrison, 1987) contains words which are used frequently but only specialists in astro-physics can fully the concepts of this phrase. This factor seems to relate to background knowledge which is one of the reader factors. It is difficult to assess it because it is a personal factor. We hardly have ready criteria for judging it reliably.

Syntax refers to “the rules about how words are arranged and connected to make phrases and sentences” (Macmillan Dictionary, www, 2009). Most teachers believe that the more complex a text is in terms of syntax, the harder it will be to comprehend. Harrison (1987) presents five types of difficulty related to syntax. These types are: 1) active versus passive verbs (active verbs are easier to read and to recall than passive verbs, and they are less likely to be misunderstood when a negative statement is made); 2) nominalization versus active verb (active verbs are easier to comprehend and to recall than an abstract noun formed from the verb); 3) modal verbs (modal verbs such as might, could, may and should cause comprehension difficulties for poor readers, and make recall more difficult for fluent readers); 4) clauses per sentence (the more clauses there are in a sentence, the more difficult it is to understand); and 5) compression and substitution (compression reduces sentence length but can make comprehension more difficult). The idea of active versus passive verbs is used in the calculation of passive sentences readability scores and the measure of syntactic complexity by clauses per sentence is employed in several readability formulas like Homan-Hewitt Readability Formula. However, it is uncertain that these are reliable indications of structural and text difficulty. This is because as mentioned in 1.2.1, simplifying the syntax by splitting complex sentences into component clauses does not always make texts easier (Davison and Kantor, 1982). Moreover, these syntactic variables do not always act independently within a sentence. They interact, often in ways which will be extremely difficult to assess experimentally (Harrison, 1987). Hence, syntax is not included in the development of the index.

Although there are several factors affecting text difficulty, most of them such as motivation, interest, background knowledge, conceptual difficulty, seem to be imponderables. Some of them like syntax are unlikely to be a reliable indicator of text difficulty (see 1.2.1). Some of them, namely, legibility of prints, illustration and color tend to rely on readers' preferences and how they affect text difficulty is still questionable so it seems to be difficult to draw conclusions about legibility of prints, illustration and color which will help teachers and students make decisions about difficulty of texts. Therefore, vocabulary difficulty tends to be worth studying as an indicator of text difficulty because of its strong relationship with reading comprehension (see 2.5).

According to Chall (1958); Klare (1974-1975); Anderson and Freebody (1981); Stahl (1986); Laufer (1992, 1997); Qian (2002); Adams (2004) and Zhang and Annual (2008), vocabulary difficulty has consistently been found to be the most significant predictor of readability of a text. It appears to contribute to several traditional readability prediction studies (Davey, 1988) such as Dale and Chall readability formula, Spache readability formula, Fog index, Flesch readability formula, and Flesch-Kincaid readability test. According to Nation and Coady (1988), vocabulary difficulty has often been identified by familiarity of vocabulary and word length. Some of the traditional formulas such as Fog index, Flesch readability formula, and Flesch-Kincaid readability test use word length to identify difficult words because it is claimed that longer words tends to be more difficult than shorter ones. Some of the readability formulas like Dale and Chall readability formula and Spache readability formula employ vocabulary familiarity to identify vocabulary difficulty because they believe that a sentence with high frequency words is more readable than a sentence with low frequency words.

In the present study, vocabulary familiarity is also used to identify vocabulary difficulty. It is categorized by frequency of words occurring in everyday life. On the one hand, when students see a high frequency word or a word that frequently occurs in everyday life, the students tend to be familiar with the word and it is the most likely that they would find the word easy. As a result, text with a lot of high frequency words tends to be easy. On the other hand, when students see a low frequency word or a word that does not frequently occur in everyday life, the students may never see the word or may find the word unfamiliar so it is possible that they will find the word difficult. When a text has a lot of low frequency words, it tends to be more difficult.

We refer to the proportion of low and high frequency words in a text as Lexical Frequency Profile (LFP). Several researchers such as Astika (1993); Meara (1993); Laufer (1994); Laufer and Nation (1995); Goodfellow, Jones and Lamy (www, 2002); Muncie (2002); Meara (2005) make use of LFP in their studies. The uses of LFP in these studies can be categorized into two areas. One is to use LFP to evaluate words that students would be exposed to and the other one aims to analyse LFP in order to evaluate words used in students' written work.

At present, LFP is also used to evaluate text difficulty for various levels of students (Cobb, 2007). This evaluation is based on the findings presented by Liu and Nation (1985) and Laufer (1989). They found out that if a student knows 95% of tokens or running words in a text, that student can successfully guess unknown words and comprehend the text. If a student is, for example, strong at the 0-1,000 level and weak at all levels beyond that, then it will be good for him or her to read texts that present about 95% of words from the level that he/she knows and the other 5% from the levels beyond his/her level of vocabulary knowledge (Cobb, 2007). This approach of evaluating texts has been applied to a number of sources like Valcourt and Wells (1999) which present evidence to show that the approach works well in practice (Meara, 2005). However, the research on the investigation of LFP as a method of indicating text difficulty seems to be rare. Therefore, this study aims to investigate whether LFP can be used to indicate difficulty of texts.

2.5.2 Prediction of Reading Comprehension

One of the questions concerning receptive vocabulary and reading comprehension that has provoked interest is how many words in a foreign language students should know in order to be able to read an authentic text (Laufer, 1989) because "knowing students' receptive vocabulary size provides teachers with a gauge as to whether those students will be able to comprehend a text" (Webb, 2008). As a result, it is worth studying vocabulary threshold which marks the boundary between not having and having enough vocabulary knowledge for comprehending an authentic or unsimplified text because if a student has not crossed the threshold, then sufficient comprehension is not possible or the chances of comprehending adequately are low. If the reader has crossed the threshold, then, comprehension is possible (Nation, 2001).

Studies on the vocabulary threshold are such as Laufer (1989); Hirsh and Nation (1992); Laufer (1992); Hu and Nation (2000); Chujo and Utiyama (2005); Nation (2006) suggest that there are certain vocabulary thresholds that determine whether students or readers will be able to successfully read a text. For example, Laufer (1989) reports that students or readers who know 95% of tokens or running words in a text can successfully guess unknown words and comprehend the text. Laufer (1992) find out that the receptive knowledge of the first 3000 most frequent word families, at a minimum, is necessary for students to understand unsimplified texts and Hirsh and Nation (1992) suggest that knowing at least 5000 most frequent word families is required for reading to be enjoyable.

From the examples of acceptable thresholds, we can see that the aforementioned studies on the vocabulary thresholds focus on two aspects of relationship between vocabulary and reading comprehension which are 1) how many words readers should know so as to understand authentic texts and 2) how many words in a text that readers should know in order to comprehend that text.

These two aspects of relationship associate with two out of three aforementioned components of the vocabulary-based readability index. These two components are: 1) prediction of reading comprehension at different vocabulary size; and 2) prediction of reading comprehension at different TSVK.

2.6 Summary

In this chapter, we have seen that receptive vocabulary closely relates to reading comprehension. There are several attempts to use this relationship to predict readability. However, most of them tend to employ only word length or familiarity of words to identify vocabulary difficulty. None of them integrates both difficulty of vocabulary in texts and readers' vocabulary knowledge into the prediction. This integration seems to be a promising approach in predicting text difficulty and reading comprehension. So far, we have seen the need to devise the index and the literature related to the development of the index. We will look at research methodology of the present study in Chapter 3.

CHAPTER 3

RESEARCH METHODOLOGY

The main purpose of this study was to devise the vocabulary-based readability index and investigate the extent that the vocabulary-based readability index would be valid. Before the vocabulary-based readability index could be devised, it was crucial to pilot some of the research instruments especially a yes/no test and a reading comprehension test. As a result, there were four main phases of investigation: 1) pilot work of the yes/no test; 2) pilot work of the reading comprehension test; 3) development of the vocabulary-based readability index and 4) validation of the vocabulary-based readability index. Overview information of research methodology including participants, instruments, data types and data analysis of each of the four phases is presented in Table 3.1 below.

Table 3.1: Overview Information of Research Methodology

Phase	Participants	Instruments	Data Types	Data Analysis
Phase 1: pilot work of yes/no test	75 students	- Yes/no test - Translation test	Quantitative	- Cronbach's Alpha - Pearson *r*
	62 students	- Yes/no test - Vocabulary Knowledge Scale (VKS)	Quantitative	- Frequency - Percentage
Phase 2: pilot work of reading comprehensio n test	5 lecturers	Questionnaire for face validity of reading comprehension test with four passages	Quantitative	Mean
	90 students	Reading comprehension test with four passages	Quantitative	- Cronbach's Alpha - Item analysis

Table 3.1 (cont.): Overview Information of Research Methodology

Phase	Participants	Instruments	Data Types	Data Analysis
Phase 3: development of vocabulary-based readability index	102 students	- RANGE - Yes/no test - Self-report on unknown words for testing TSVK - Translation test for testing TSVK - Reading comprehension test with four passages	Quantitative	- Descriptive statistics (mean and standard deviation) - Three-way ANOVA - Scheffé Test - Pearson *r* - Simple regression analysis
Phase 4: validation of vocabulary-based readability index	- 3 students - 3 lecturers	Questionnaire for investigating face validity of the vocabulary-based readability index	Quantitative	Frequency
	- 45 students - 5 lecturers	Questionnaire for investigating concurrent validity of LFP as an indicator of text difficulty	Quantitative	Spearman's rho
	80 students	- RANGE - Yes/no test - Self-report on unknown words for testing TSVK in the reading comprehension test with five passages from TOEIC tests - Reading comprehension test with five passages from TOEIC tests - Translation test for testing TSVK in the reading comprehension test with five passages from TOEIC tests	Quantitative	Pearson *r*
	30 students	- RANGE - Yes/no test - Self-report on unknown words for testing TSVK in the reading comprehension test with four passages - Reading Comprehension Test with four passages - Translation test for	Quantitative	Pearson *r*

Table 3.1 (cont.): Overview Information of Research Methodology

Phase	Participants	Instruments	Data Types	Data Analysis
		testing TSVK in the reading comprehension test with four passages - Questionnaire for investigating predictive validity of the prediction of reading comprehension		

From Table 3.1, we can see that the type of all of the data was quantitative because the data were in numerical form. We can also see that there were four main phases: 1) the pilot work of the yes/no test; 2) the pilot work of the reading comprehension test; 3) the development of the vocabulary-based readability index and 4) the validation of the vocabulary-based readability index. The details of research methodology were presented in four main parts based on the four phases (Phase 1, Phase 2, Phase 3 and Phase 4) as follows.

3.1 Phase 1: Pilot Work of the Yes/no Test

This phase was to investigate reliability, concurrent validity and construct validity of the yes/no test in order to find out whether the yes/no test was appropriate for the actual study. This section aims to present research methodology including participants, instruments, data collection procedures and data analysis in the pilot work of the yes/no test.

3.1.1 Participants in the Pilot Work of the Yes/no Test

From Table 3.1, there were two groups of participants taking part in the pilot work of the yes/no test. Each of the two groups consisted of 75 and 62 Thai university students respectively. The information about each group is presented below:

The participants in the first group were 75 Thai university students. The name of the university cannot be mentioned in this study because the researcher was asked to keep the name anonymous. They were first year participants studying English II course. They were about 18-20 years old when the data were collected. They majored in science-oriented fields. The participants took two English classes per week in reading and writing taught by the same teacher. They studied English as a foreign language because English was not the language used in their community (Cook, 1991). They had studied English since they were grade one students so they had at least twelve years of exposure to English. They had a wide range of levels of English language proficiency from beginners to pre-intermediate. In the pilot work, they were named Participants A1-A75.

The participants in the second group were 62 Thai second year students at a public university in Thailand. Their ages ranged from 19-21 years of age. They shared similar educational backgrounds, studying English as a foreign language, having approximately 13-16 years of exposure to English and majoring in Chemistry. They had a wide range of levels of English language proficiency from beginners to pre-intermediate. The data were collected when the participants were taking a Fundamental English course. The aim of the course was to improve the students' English language proficiency in four skills: reading, writing, listening and speaking. They learned new vocabulary both inside and outside the classroom through the reading materials assigned by the teacher. The participants in this group were given identification codes B1-B62.

After collecting the data, four out of 62 participants in the second group (B1, B5, B13 and B17) were discarded from the pilot work of the yes/no test. Two of them (B5 and B17) were ignored because they did not complete the yes/no test. The researcher first assumed that they did not know any words in the test. Surprisingly, they could correctly translate 11 and 13 words on the VKS respectively. The other two (Participants B1 and B13) were discarded from the study because they did not complete every item in the VKS so we could not know the category that represents their vocabulary knowledge. Therefore, data from 58 participants would be analyzed in the pilot work.

3.1.2 Instruments for the Pilot Work of the Yes/no Test

In order to be able to investigate reliability, concurrent validity and construct validity of the yes/no test, three main research instruments were used. These instruments were a yes/no test, a translation test and Vocabulary Knowledge Scale (VKS). Each instrument is presented in detail below.

3.1.2.1 Yes/no Test in the Pilot Work of the Yes/no Test

The yes/no test was used to test ability to recognise target words and understand the meaning (Beeckmans, Eyckmans, Janssens, Dufranne and Van de Valde, 2001). It is a simple test format (Anderson & Freebody 1983; Read 2000) that is easy to construct, administer and score (Eyckmans, www, 2004). It was used because it permits a large number of words to be tested in a limited amount of time (Meara and Jones 1988).

The yes/no test in this study consisted of four different levels. These levels were named Level 1, Level 2, Level 3 and Level 4 in the yes/no test. Each level consisted of 50 words. Thirty of them were target words that were randomly selected from individual four word lists based on the first, the second, the third and the fourth 1000 most frequent word families from British National Corpus (BNC). Level 1 consisted of 30 target words randomly selected from the first 1000 most frequent word families based on the BNC. Level 2 was comprised of 30 target words from the second 1000 most frequent word families and so on. The other 20 words were pseudowords or words that do not exist. They were constructed by Meara and his colleagues at University of Wales Swansea (www, 1992). The instructions of the yes/no test was adapted from Meara and Jones (1988) and Meara (www, 1992) and translated into Thai. The yes/no test used in the study is shown in Appendix A. There were two forms of tests (Test A and Test B) differing in order so as to eliminate possibility of cheating.

After the yes/no test had been completed, it was crucial to score it. According to Huibregtse, Admiraal and Meara (2002), there are four possible scoring methods for the yes/no tests. These methods are the hit rate minus the false alarm rate (h - f), correction for guessing (cfg), Meara's Δm, and a new index based on signal detection theory. In this study, h - f was used as the scoring method of yes/no test because it is the simplest scoring method. It is easy to calculate and to explain the procedure (Huibregtse et al., 2002). Hence, any teachers who are willing to use the devised index can simply score the yes/no test by themselves.

As the name of the scoring method suggests, we need to subtract the false alarm rate (f) from the hit rate (h). To calculate the hit rate and the false alarm rate, we need to count the selected target words and the selected pseudowords in each level in the test. The number of selected target words was divided by the total number of target words (30) in each level in the test in order to calculate the hit rate. The number of selected pseudowords was divided by the total number of pseudowords (20) in each level so as to calculate the false alarm rate or f. The false alarm rate was subtracted from the hit alarm rate ($h – f$). The maximum score calculated by $h - f$ is 1 and the minimum is -1.

Although there is a considerable amount of research and development work on the yes/no test such as Meara and Buxton (1987); Meara and Jones (1988); Shillaw (1996); and Mochida and Harrington (2006) and it was found in these studies that it is likely that the test can be used without "a totally untried methodology" (Alderson and Huhta, 2005), the test was piloted again in order to make sure that the test is appropriate for Thai university students who were the participants in the study. The results of the pilot work of the yes/no test will be presented in Chapter 4.

3.1.2.2 Translation Test for the Pilot Work of the Yes/no Test

A translation test was used as an instrument in the pilot work of the yes/no test. It aimed to test recognition of written form of words and their meanings. This pilot work investigated the relationship between the participants' scores on the yes/no test and their scores on a translation test in order to investigate concurrent validity - "the extent to which a test correlates with some other test that is aimed at measuring the same skill, or with some other comparable measure of the skill being tested" (Richards and Schmidt, 2002). The translation test seems to be more appropriate to this pilot work than other measures of vocabulary because explaining the meaning of target words through translation is much easier for students than through multiple-choice items providing definitions. They do not need to respond to vocabulary items in a way that draw on English language knowledge which is not directly relevant to what is being tested (Nation, 2001). According to Eyckmans (www, 2004), it is assumed that asking the participants to provide mother-tongue equivalents of target language words was the most univocal way of testing word recognition. For word meanings, Nation (2001) affirms that the translation test is a direct way of testing whether the students can give the meaning to the target words.

The translation test for the pilot work of the yes/no test was constructed by using “Research Randomizer” (Urbaniak and Plous, www, 2007) to randomly select eight words from each of the four levels in the yes/no test so the test consisted of thirty-two target words. The selected words were finance, final, treat, local, suggest, general, reduce, responsible, invite, accuracy, fortnight, satisfy, burn, desire, indicate, tough, sympathy, manner, outcome, interfere, accuse, horror, raw, courage, classify, heritage, absent, cheerful, gender, herb, shallow, ambitious. The task in the translation test was to translate these target words into English or Thai (see Appendix B). Answer key was provided based on four dictionaries which are Oxford Advanced Learner’s Dictionary (Cowie, 1989); Collins COBUILD Dictionary (Sinclair, 1995); SE-Ed’s Modern English-Thai Dictionary (Thiengburanathum, 1998); and English-Thai Dictionary (Sethaputra, 2003). Two of them are English-English dictionaries and the other two are English-Thai dictionaries. The participants whose answer matches the possible answers in the answer key got one mark per item while the ones, who got a wrong answer or left a blank in the test, got zero. The participants who got all correct answers got 32 marks.

Since the target words were presented in the yes/no test without any contexts, the translation test used to compare against the yes/no test for the purpose of validation should be decontextualized in order to make sure that the translation test is very likely to measure the same constructs as the yes/no test. If contextualized vocabulary tests or tests that provide cues or contexts of the target words in the test like the translation test in Nurweni and Read (1999), the constructs of these contextualized tests may be slightly different from the yes/no test because the tests presenting contexts of the target words do not test only the participants' ability to recognize target words and understanding their meanings which is the construct of the yes/no test (Beeckmans et al., 2001) but they also test the ability to guess the meanings of the words from contexts. Therefore, contexts are not provided in the translation test.

Like the yes/no test, there were two forms of the translation test (Test A and Test B). The two forms differed in order so as to eliminate possibility of cheating.

3.1.2.3 Vocabulary Knowledge Scale

The other research instrument of the pilot work of the yes/no test was Vocabulary Knowledge Scale (VKS). VKS is a five-point scale used to elicit self-perceived and demonstrated knowledge of specific written (or printed) words. The scale ratings range from unfamiliarity, through recognition of a target word and some idea of its meaning, to the ability to use the word with grammatical and semantic accuracy in a sentence (Paribakht and Wesche, 1997). Below is an example of VKS developed by Paribakht and Wesche.

Self-report Categories	
I	I don't remember having seen this word before.
II	I have seen this word before, but I don't know what it means
III	I have seen this word before, and I think it means _____________. (synonym or translation)
IV	I know this word, it means ______________. (synonym or translation)
V	I can use this word in a sentence: ____________________. (Write a sentence.) (If you do this section, please also do Section IV.)
	(Paribakht and Wesche, 1997)

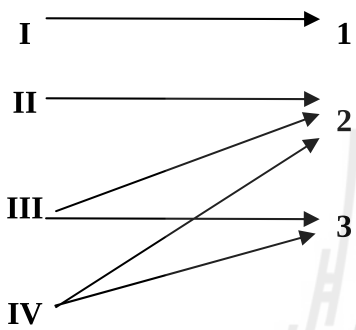
Figure 3.1: VKS elicitation scale

From the elicitation scale of VKS in Figure 3.1, we can see that the last category (V) requires the participant or test-taker to write a sentence with the target word. It helps us know whether the participant can use the target word accurately in a sentence but it was deleted when the VKS was used in this pilot work because this category is beyond the scope of the yes/no test. The yes/no test is claimed to test only the ability to recognize target words and to understand the meaning. Therefore, the VKS used in this study is a four-point scale rather than a five-point scale. Moreover, since the participants completing the VKS were all Thai, the VKS was translated into Thai in order to make sure that the VKS tests vocabulary knowledge rather than the ability to comprehend the description of each category. Before the participants were asked to complete the test, a Thai lecturer was asked to check the clarity and appropriateness of the language used in the test.

Eight words were randomly selected from each level in the yes/no test by using "Research Randomizer" (Urbaniak and Plous, www, 2007). Since there were four levels in the yes/no test, there were 32 words in the VKS: quiet, garden, local, health, burn, invite, ordinary, drag, soul, extract, manner, merry, jungle, grove, elegant, spontaneous, brilliant, cheap, couple, community, nasty, grace, tough, holy, donate, elbow, greedy, trophy, ripe, heave, legislate and disruption. The Thai and the English versions of the VKS are presented in Appendix C.

Although the number of points in the scale is different from the scales developed by Paribakht and Wesche, the scoring method is the same. A participant who reports that he/she has never seen the target word by selecting category I will get one mark. The participant will get two marks if they select category II or if they select category III or IV but they cannot write any correct synonym or meaning. A score of 3 indicates that an appropriate synonym or translation has been given for categories III and IV. By way of illustration, the possible scores for a target word on the VKS and their relationship to the self-report categories are given in Figure 3.2.

Self-report Categories **Possible Scores** **Meaning of Scores**



The word is not familiar at all.

The word is familiar but its meaning is not known.

A correct synonym or translation is given.

Adapted from Paribakht and Wesche (1997)

Figure 3.2: VKS scoring categories and meaning of scores

From Figure 3.2, we can see that the minimum possible score is one and the maximum is three. The participant who does not know anything about the word will get a score of one and the participant who knows the meaning of the target word will get a score of three.

3.1.3 Data Collection Procedures in the Pilot Work of the Yes/no Test

Before the data were collected, the participants were informed the purpose of the pilot work. They were also given the opportunity to not participate or to withdraw at any time they wished. All of them agreed to take part in the pilot work and none chose to withdraw. They were asked to sit on every other chair in order to promote examination conditions and avoid peeking.

When the yes/no test was administered, all of the participants in the pilot work were asked to circle the words in the test that they know well enough to say what they mean. They were also informed that the test contains pseudowords or non-existing words but they do not have any clues of how many and their location in the test (Huibregtse et al., 2002).

As mentioned in 3.1, the pilot work of the yes/no test aimed to investigate reliability, concurrent validity and construct validity of the yes/no test. In order to investigate reliability and concurrent validity of the yes/no test, 75 Thai university students were asked to complete the yes/no test and then the translation test for the pilot work of the yes/no test. They were not allowed to use any dictionaries. They had 45 minutes to complete the yes/no test and the translation test.

One month later, 62 Thai university students taking part as the participants in the pilot work were asked to complete the yes/no test and the VKS respectively in order to investigate the construct validity of the yes/no test. Before the yes/no test and the VKS were administered, these participants were told the purpose of the pilot work, the formats and the instructions of the yes/no test and the VKS because these two tests were quite new to them. They had 45 minutes to complete the yes/no test and the VKS. Dictionaries were not allowed.

3.1.4 Data Analysis for the Pilot Work of the Yes/no Test

In order to investigate reliability, concurrent validity and construct validity of the yes/no test, Cronbach's alpha, Pearson product-moment correlation coefficient or Pearson *r*, frequency and percentage were used as methods of data analysis. Each method is presented in detail below.

3.1.4.1 Cronbach's Alpha for the Pilot Work of the Yes/no Test

Cronbach's alpha (α) is a measure of the degree to which the items are homogenous or consistent with each other (Richards and Schmidt 2002). It was used in this pilot work to investigate reliability of the yes/no test. In order to calculate Cronbach's alpha coefficients in SPSS 16.0, the answers of the participants in the first group were entered or typed into SPSS. Numbers 1 to 4 were used to represent the participants' answers on the yes/not test in the SPSS. Number 1 referred to the "yes" response to a target word or the selection of a target word, number 2 referred to the "no" response to a target word, number 3 referred to the "yes" response to a pseudoword and number 4 referred to the "no" response to a pseudoword. In order to analyze the data, the reliability analysis command in SPSS was Analyze, Scale and then Reliability. The result of Cronbach's alpha was in terms of a coefficient between 0 and 1. The test was said to be reliable if the Cronbach's alpha coefficient was higher than the acceptable criterion of 0.70 (Fraenkel and Wallen, 1993).

3.1.4.2 Pearson Product-Moment Correlation for the Pilot Work of the Yes/no Test

Pearson product-moment correlation or Pearson *r* is a measure of the correlation between two variables giving a correlation coefficient ranging from +1 to -1. It was employed in this pilot work to investigate concurrent validity of the yes/no test. It calculated correlation coefficients between scores on the two forms of the yes/no test and the two forms of the translation test with the assistance of SPSS 16.0. To be able to analyze the data, the command in SPSS was Analyze, Correlate and then Bivariate.

3.1.4.3 Frequency and Percentage for the Pilot Work of the Yes/no Test

Frequency was used in this pilot work to present the numbers of responses where performance on the yes/no test and the VKS matched each of the four conditions. These conditions were: 1) the participants circled a target word in order to report that they know the meaning of the word and when they were asked to complete the VKS, they could translate the word correctly (Condition A); 2) the participants reported that they knew the meaning of the target word but they could not give accurate meaning or synonym of the target word (Condition B); 3) the participants reported in the yes/no test that they did not know the meaning of the word and they could not give accurate translation in the VKS (Condition C); and 4) the participants did not say yes to the target word but they could translate the word correctly (Condition D).

After counting the responses where performance on the yes/no test and the VKS matched each of the four conditions, the numbers of responses were converted into percentage. Percentage is a way of expressing a number as a fraction of 100. It was denoted in the pilot work of the yes/no test using the percent sign "%". In order to calculate percentage, the total number of responses for each condition was divided by the total amount of participants and then multiplied by 100.

3.2 Phase 2: Pilot Work of the Reading Comprehension Test

This phase was to find out whether the reading comprehension test was appropriate for the study. The reading comprehension test was investigated in terms of face validity, reliability, item difficulty and item discrimination. This section describes the research methodology of the pilot work of the reading comprehension test. It includes the information on participants, instruments, data collection procedures and data analysis.

3.2.1 Participants in the Pilot Work of the Reading Comprehension Test

According to Table 3.1, two groups of participants took part in the pilot work of the reading comprehension test. The participants in the first group were 5 university lecturers and the participants in the second group were 90 university students. The 5 experienced university lecturers who were supposed to be familiar with teaching and testing reading were asked to take part in the investigation of face validity of the reading comprehension test so as to check whether the test appears to test the ability to scan and paraphrase specific information or not. In addition, the 90 participants who were university students were asked to take part in the investigation of reliability and item analysis in order to check the extent to which the test gave consistent results and to find out whether the individual items in the test were appropriate for university students. The details of each group are presented below.

Five Thai university lecturers were asked to take part in the pilot work of reading comprehension test. They had at least three years of experience of working as a lecturer in a university. They had experienced teaching fundamental English courses at a public university and constructing tests of English language skills including reading, writing, speaking and listening. Their ages ranged from 30 to 45 years of age. They were named Teacher A, Teacher B, Teacher C, Teacher D and Teacher E.

The participants in the second group in the pilot work of the reading comprehension test were 90 Thai university students. Forty-five of them were studying the last class of a fundamental English course called English IV at Suranaree University of Technology. The course aims to improve the students' reading skills. They had two classes per week taught by the same teacher. The other 45 participants were students at Kasetsart University. They were studying the second class of the fifth fundamental English course intending to improve students' reading proficiency. They had a three-hour class per week. They had studied four fundamental English courses at the university. They were about 21-23 years old. They shared similar educational backgrounds, studying English as a foreign language, having approximately 16-19 years of exposure to English and majoring in science-oriented fields. They had a wide range of levels of English language proficiency from beginners to pre-intermediate. They learned new vocabulary both inside and outside the classroom through the reading materials assigned by the teacher.

3.2.2 Instruments for the Pilot Work of the Reading Comprehension Test

There were two research instruments in the pilot work of the reading comprehension test. These instruments were 1) a reading comprehension test and 2) a questionnaire for investigating face validity of the test. Each instrument is presented below.

3.2.2.1 Reading Comprehension Test in the Pilot Work of the Reading Comprehension Test

The reading comprehension test aimed to test the ability to scan and paraphrase specific information. There were four sections in the reading comprehension test. Each section was comprised of a reading passage and six comprehension questions so there were four reading passages and twenty-four questions in the test. The passages were 130-160 tokens long. The reading comprehension test is given in Appendix D. After the reading comprehension test had been completed, it was scored. Each of the four sections was scored separately. Each test item was worth one mark. The participant who got a correct answer on an item got one mark. The participant whose answer did not match the answer key got zero. The maximum score on each section was six and the minimum was zero.

According to Chapter 2, there are several factors influencing reading comprehension. These factors are, for example, interest and motivation, background knowledge, legibility of print, illustration and colour, conceptual difficulty, LFP and syntax. Some of these factors such as legibility of prints, illustration, color, LFP and syntax can be controlled while some factors such as interest, motivation, background knowledge, conceptual difficulty cannot be controlled. These controllable factors were controlled in this study in order to ensure that LFP is likely to be the only factor influencing text difficulty and reading comprehension. In order to control the factors, guidelines for controlling these variables and selecting reading passages were constructed and presented below.

3.2.2.1.1 Text Type

Researchers have studied how different text types might influence comprehension in L1 studies (Hiebert, Englert, and Brennan, 1983; Englert and Hiebert, 1984; Mandler, 1978; Abu-Rabia, 1998; Shin, 2002), L2 and foreign language studies (Carrell, 1984, 1985: Okumura, 1998; Ayatollahi, 2003; Kobayashi, 2004; Parvaresh and Nemati, 2008) and both L1 and L2 studies (Brantmeier, 2005). Dennis (1982) and Brantmeier (2006) report that text type is likely to be a factor that influences how a reader reads, comprehends and/or remembers the content. Hence, it is necessary for us to control it. Werlich (1976 quoted in Trosborg, 1997) classifies texts into five types: description, narration, exposition, argumentation and instruction.

1) Description refers to texts that show differentiation and interrelation of perceptions in space.

2) Narration refers to texts that show differentiation and interrelation of perceptions in time.

3) Exposition refers to texts that provide information or general concepts about and explain a particular subject.

4) Argumentation refers to texts that evaluate relations between concepts through the extraction of similarities, contrasts, and transformations.

5) Instruction refers to texts that provide planning of future behavior with option (advertisements, manuals, recipes) and without option (legislation, contracts)

To be able to select text type of reading passages in the reading comprehension test, text types of reading passages in entrance examination and reading practice books were investigated. After looking through reading passages in entrance examination in academic years 1992-2003, some commercial books and practice books such as Multiple Reading Skills (Boning, 1995); Cambridge Practice

Tests for IELTS 1 (Jakeman and McDowell, 1996); IELTS Preparation and Practice (Sahanaya, Lindeck and Stewart, 1998); 101 Helpful Hints for IELTS Academic Module (Adams and Peck, 2000); Cambridge IELTS 2 (University of Cambridge Local Examinations Syndicate, 2000); Cambridge IELTS 3 (University of Cambridge Local Examination Syndicate, 2002); Insight into IELTS Extra with Answers (Jakeman and McDowell, 2003); 404 Essential Tests for IELTS (Scovell, Pastellas and Knobel, 2004); Barron's TOEFL iBT 2006-2007 (Sharpe, 2006); and The Official Guide to the New TOEFL iBT (Educational Testing Service, 2007c), it was found out that expository texts are frequently used. In addition, when we look through research concerning reading and text types, we can see that the most prominent investigated text type is exposition (Ogle, 1986; Salager-Meyer, 1991; Varnhagen, 1991; Rosa, 1994; Lehto, Scheinin, Kupiainen and Hautamaki, 2001; Degand and Sanders, 2002; Linderholm and Van Den Broek, 2002; Kobayashi 2004; Kobayashi, 2007; Ozgungor and Guthrie, 2004; Fang, 2008) It is sometimes examined with narration (Okumura (1998); DuBravac and Dalle (2002); Irene-Anna, Polyxeni, Christina and Panayiota (2005).

When students can prominently find expository texts in the reading sections of Entrance Examination, IELTS as well as TOEFL, it was selected for two reasons. One is that the students tend to be familiar with this text type. The other reason is that when students tend to regularly find this text type, it is likely that it is the text type of reading passages that most of the teachers and the students would like to select. It will then be good if the vocabulary-based readability index was devised in order to help the teachers or students select reading passages with the text type that the students are prominently exposed to. In addition, as mentioned earlier, this text type is investigated in many pieces of research so it seems to play an important role in reading. Therefore, exposition was selected as the text type of the reading passages in the reading comprehension test. To be able to select expository texts or reading passages with exposition, we need the guidelines below:

Guideline 1: Fiction texts are ignored.

Guideline 2: Expository passages are subject oriented. They focus on a specific topic.

Guideline 3: Reading passages that give examples; describe a process of doing or making something; analyse causes and effects; define a term or concept; divide something into parts; classify something into categories (Richards and Schmidt, 2002); inform, explain or enumerate something; convey an idea; or demonstrate knowledge of a subject matter are expository texts.

After analysing text types of reading passages, we had a number of expository passages. We, then, looked at another aspect of reading needed to be controlled – length of the reading passages.

3.2.2.1.2 Text Length

Few researchers have commented on the effects of long texts on readability and a study on the impact of text length on students' reading comprehension does not show a significant impact on the students' performance on reading tests with different lengths of reading passages (Mehrpour and Riazi, 2004). However, it has an effect on the participants' motivation (Huhmann, Mothersbaugh, and Franke, 2002; Mori, 2002) to complete a test. Motivation is a factor relating to success in second or foreign language learning (Gardner and Lambert, 1972; Simon, Howe and Kirschenbaum, 1972, Wen, 1997; Melendy, 2008). When the motivation is low, students' English language proficiency or performance tends to decline (Srivarakan and Tananuraksakul, 2002; Bolanle, 2007). Therefore, it is crucial to select reading passages with the same length in order to prevent any effect that might be caused by text length. This study is not the first one that controls the text length. Some studies in readability like Ownby (2005) also controls numbers of words of the sample of texts used in the studies in order to prevent any effect of the length of texts on readability.

As mentioned earlier that the reading passages in the reading comprehension test were 130-160 tokens in length and there were only six test items in each section because it was estimated that this was the optimal length for university students. After the ninety participants who were 45 participants from Suranaree University of Technology and 45 participants from Kasetsart University completed the test, ten of them (five from each university) were randomly selected and asked to rate the suitability of the text length from too short (1) to too long (5). They all selected number 3 which meant that the length was optimal. This shows that the reading passages with 130-160 tokens tend to be suitable for these participants who were assumed to have similar reading proficiency and educational backgrounds to the participants in Phase 3: the development of the vocabulary-based readability index and Phase 4: the validation of the vocabulary-based readability index.

The following guideline used to select reading passages with optimal length and the length of each reading passage selected for the reading comprehension test are shown below:

Guideline: The reading passages, which are longer or shorter than 130-160 tokens or running words counted by Microsoft Word 2003, are discarded.

Table 3.2: Length of Reading Passages in the Reading Comprehension Test with Four Passages

	Passage I	**Passage II**	**Passage III**	**Passage IV**
No. of words	145	151	130	146

After getting expository texts with 130-160 tokens, LFP of these texts was analysed following the guideline for analyzing LFP (see below).

3.2.2.1.3 Lexical Frequency Profile of Reading Passages in the Reading Comprehension Test in the Pilot Work of the Reading Comprehension Test

As mentioned in Chapter 2, LFP refers to proportion of low and high frequency words. A reading passage with a lot of high frequency words tends to be easier than a passage with a lot of low frequency words because readers including the participants tend to be more familiar with the high frequency words than low frequency ones. Hence, controlling LFP enables us to classify the level of difficulty of reading passages in order that we can investigate whether the participants find the four reading passages in the reading comprehension test varied in terms of difficulty when LFP is varied. The control of LFP is based on the assumption that readers who know 95% of tokens in a reading passage can understand that passage (Laufer, 1989). The guideline for selecting reading passages with suitable LFP is given below:

Guideline: Expository texts with 130-160 tokens that have lower than 95% of tokens from the first word list; the first and the second; the first, the second and the third; word lists or all of the four word lists are discarded.

According to the guideline, Passage One or the first passage in the test should have 95% of tokens from the first frequency band or the first 1000 most frequent word families from the BNC, Passage Two has 95% of words from the first and the second 1000 most frequent word families and so on. The numbers and percentage of LFP from the four selected reading passages are presented Table 3.3 below:

Table 3.3: Lexical Frequency Profile of the Four Reading Passages in the Reading Comprehension Test in the Pilot Work of Reading Comprehension Test

	Passage One	**Passage Two**	**Passage Three**	**Passage Four**
Tokens/% in the first BNC word list	**139/95.86**	**129/84.87**	**110/84.62**	**110/78.01**
Tokens/% in the second BNC word list	6/4.14	**16/10.53**	**10/7.69**	**10/7.09**
Tokens/% in the third BNC word list	0/0.00	2/1.32	**7/5.28**	**6/4.26**
Tokens/% in the fourth BNC word list	0/0.00	0/0.00	0/0.00	**10/7.09**
Off-list	0/0.00	5/3.29	3/2.31	5/3.55

The numbers and the percentage of tokens from each of the four BNC word lists are given in Table 3.3. The number before the slash (/) refers to numbers of tokens from each BNC word list and the number after the slash (/) refers the percentage of tokens from each BNC word list. By way of illustration, let us look at LFP in Passage One presented in the above table. It is shown that 139 tokens out of 145 tokens or running words in Passage One are from the first BNC word list consisting of the first 1000 most frequent word families and these 139 tokens make up 95.86% of the total tokens in the text.

We can see from Table 3.3 that all of the selected passages meet the guideline for LFP. Passage One has 95.86% of tokens from the first BNC word list; 95.40% of tokens in Passage Two are from the first two word lists (the first and the second 1000 most frequent word families); 96.92% of tokens in Passage Three are from the first three word lists (the first, the second and the third 1000 most frequent word families) and 96.45% of tokens from Passage Four are from the first four word lists (the first, the second, the third and the fourth 1000 most frequent word families).

3.2.2.1.4 Sentence Length

Some traditional readability formulas like Flesch reading formula, Dale-Chall readability formula and Flesch-Kincaid readability test use average sentence length (total words divided by total sentences) as a proxy of grammatical complexity in order to predict readability (Dale and Chall, 1948; Flesch, 1979; Ownby, 2005). It is normally considered that long sentences are difficult to understand because they often contain a number of subordinate clauses which may cause grammatical complexity. Moreover, when readers read long sentences, it is difficult for them to bear all of their points in mind because there are often so many points in a long sentence. It is hard for the readers to remember the first part of the sentence when they are reading the last part. In other words, long sentences overload the memory system while short sentences do not. However, it is not definitely true that passages written in short sentences will always be better understood. Chapanis

(1965; 1988 quoted in Hartley, 1996) provides abundant examples of short texts that are difficult to understand. Below is one of the examples:

PLEASE
WALK UP ONE FLOOR
WALK DOWN TWO FLOORS
FOR IMPROVED ELEVATOR SERVICE

The above notice can be interpreted into several ways, for instance, "to get on the elevator I must either walk up one floor, or go down two floors" and "to get on the elevator I must first walk up one floor and then down two floors". Actually, this notice means "Please, don't use the elevator if you are only going a short distance". This example shows that short sentences are not always easy to understand.

The effect of the average sentence length on readability of reading passages is still questionable. However, there is no research stating that the sentence length which may affect text complexity does not make texts difficult at all. Therefore, it is good to control the average sentence length of the reading passages used in the study in order to reduce the possibility that the variable might distort the data. The guidelines are presented below:

Guideline 1: The average sentence length which is not greater than 20 to 30 words is considered acceptable (Hartley, 1996).

Guideline 2: The average sentence length of the four reading passages should be similar.

The average sentence length of the four reading passages presented in Table 3.4 below was calculated by using the spelling and grammar tool of Microsoft Word 2003. When the spelling and grammar check is completed, the average sentence length or words per sentence is presented on the window of readability statistics.

Table 3.4: The Average Sentence Length of the Four Reading Passages in the Reading Comprehension Test

	Passage One	Passage Two	Passage Three	Passage Four
Average Sentence Length	18.1	18.8	18.5	18.2

From Table 3.4, we can see that the four reading passages have similar average sentence length. The average sentence length of the four reading passages is 18s which is fewer than 20. This means that the average sentence length of the passages is satisfactory because it meets the above guidelines.

So far, we have seen how some factors that might influence reading comprehension were controlled. It is interesting to see how the reading comprehension test was constructed. The construction of the test items in the reading comprehension test was based on the test specifications below:

Table 3.5: Test Specifications of the Reading Comprehension Test

Purpose of the test	To test the ability to scan and paraphrase or restate specific information
Test-takers	Thai male and female university students. They study English as a foreign language. They are about 17-23 years old.
Structure of the test	There are four sections in the test. Each section is comprised of a reading passage, six test items or reading comprehension questions. This test is not related to any particular course of study of any students including the participants taking part in the study.
Test method	Multiple-choice questions are used in the test. Each item is made up of three parts: 1) the question or stem; 2) a single correct answer; and 3) three distracters or plausible answers
Text type	Exposition
Text length	130-160 words
Level of cognitive domain	Knowledge and comprehension
Level of comprehension	Literal comprehension
Criteria for scoring method	The answer key was provided for markers. The test-takers or participants who choose a correct answer get one mark per test item and the ones who choose distracters get zero.

According to Chapter 2, reading comprehension is defined as the level or degree of readers' understanding of a text and the understanding of the text is defined as the ability to scan or look for specific information and paraphrase it so the purpose of the reading comprehension test of the present study is to test the participants' ability to scan and paraphrase specific information.

In order to test the participants' ability to scan and paraphrase specific information, the traditional approach in measuring reading comprehension was selected. This approach is to write test items that ask questions about specific texts (Pyrczak, 1972) in order to ask the participants to scan and paraphrase specific information. A variety of test methods has been employed by test writers to assess the ability to scan and paraphrase (Perkins, 1984). These methods are, for example, multiple choice questions, matching techniques, ordering tasks, dichotomous items, short-answer tests (Alderson, 2000). There have been several attempts such as Shahomy (1984); and Kobayashi (2002) confirming that different test methods would yield different results. In order to make sure that test methods would not affect the participants' reading proficiency or performance on the reading comprehension test, only one test method was used in the test and the selected method was multiple-choice technique. It was chosen for four reasons: 1) this test method is one of the most commonly used types (Heaton, 1975) for testing reading comprehension (Alderson, 2000) especially scanning and paraphrasing so all of the participants tend to be familiar with it; 2) multiple-choice questions allow objective scoring (Bensoussan, 1984; Seaman, www, 2003) so it is easy to mark (Madsen, 1983; Higgins and Tatham, www, 2003); 3) it limits assessment bias caused by participants' or test-takers' poor writing skills (Mandernach, www, 2003); and 4) the selection of multiple-choice test format was influenced by comprehension tests of McCall and Crabbs (1925). There are six books in the series labeled A-F. Each book provides 60 reading passages followed by 8 multiple-choice questions. Questions are set up like standardized test questions so as to familiarize children with the format (Cathy Duffy Reviews, www, 2009). These tests were used most widely as a basis for deriving readability formulas. Chall (1958) called the McCall-Grabbs passages the best criteria devised for readability work because the passages are carefully graded in order of difficulty in terms of linguistic variables such as word frequency and sentence length and researchers can consider the linguistic factors associated with text difficulty and combine them into a formula (Harrison, 1980). However, these tests are not used in this study for two reasons: 1) this study does not apply the linguistic factors or variables mentioned in the books and 2) the books are unfortunately not available in Thailand. However, it is a good idea to familiarize the students by using the multiple-choice questions. Hence, the multiple-choice technique is used in the reading comprehension test of the present study.

According to Bloom, Engelhart, Furst, Hill and Krathwohl (1956), there are six levels of cognitive domain which require test-takers or participants to demonstrate 1) knowledge i.e. the student remembers information previously read and recalls; 2) comprehension i.e. the student understands what is being read, interprets and restates into own words; 3) application i.e. the student applies information or skill to reach an answer; 4) analysis i.e. the student breaks down the information into its constituent elements or parts so that the organizational structure may be understood; 5) synthesis i.e. the student puts together elements or parts to form a whole; and 6) evaluation i.e. the student makes judgments about the value of material or method for a given purpose. In the present study, only knowledge and comprehension were tested because the reading comprehension test aims to only measure the ability to scan and paraphrase specific information. Application, analysis, synthesis and evaluation seem to require more complex cognition than comprehension and knowledge so they tend to go beyond the purpose of the test.

According to Chapter 2, only literal comprehension was tested because of the participants' low level of reading proficiency and because the reading comprehension test aims to test the participants' ability to scan and paraphrase specific information. Questions of interpretative and applied comprehension tend to be more difficult (Holtzman, 2008) than literal comprehension questions because they do not only require the test-takers or participants to understand the reading passages but they also involve combining their understanding with their own knowledge and intuitions (Day and Park, 2005).

The reading comprehension test was constructed based on the test specifications in Table 3.5. After constructing the test, it was given to an English native university lecturer and a Thai university lecturer in order to ask them to comment on the suitability of the stems and options of the questions or test items as well as the language use in the test and in the questionnaire. The test was then corrected following the comments.

3.2.2.2 Questionnaire for Investigating Face Validity of Reading Comprehension Test

The questionnaire for investigating face validity of the reading comprehension test was a 3-point scale. The three points on the scale were +1 (agree that the item appears to measure the objective), 0 (unsure that the item appears to measure the objective) and -1 (disagree that the item appears to measure the objective). It consisted of 24 items (see Appendix E). The purposes of this questionnaire were to investigate face validity – the degree to which the test appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of observers (Richards and Schmidt, 2002) – by asking five lecturers to rate whether they agreed that the reading comprehension test appeared to test the objective which is the ability to scan and paraphrase specific information.

3.2.3 Data Collection Procedures of the Pilot Work of the Reading Comprehension Test

When the reading comprehension test had been constructed and revised based on a Thai lecturer and a native English lecturer, it was investigated in terms of face validity in order to make sure that the test appears to test the participants' ability to scan and paraphrase specific information. The face validity of the reading comprehension test was investigated by asking five lecturers to complete the questionnaire with 3-point rating scale for investigating face validity (see Appendix E). These lecturers were informed the purposes of the reading comprehension test and the pilot work. They were asked to look through the test and complete the questionnaire without time limit. In the questionnaire, they were asked to rate whether they agreed that the reading comprehension test appeared to test the objective which is the ability to scan and paraphrase specific information on 3-point rating scale. The scale was valued +1 (agree that the item appeared to test the objective), 0 (unsure that the item appeared to test the objective) and -1 (disagree that the item appeared to test the objective). The participants were also given definitions of scanning and paraphrasing in order to make sure that all of the participants would define these terms in the same way.

After investigating face validity of the reading comprehension test, the test was distributed to 45 students at Suranaree University of Technology and 45 students at Kasetsart University on two different days in the same week. Before the test was administered, they were informed the purposes of the pilot work and the reading comprehension test. They were also notified their rights to withdraw at anytime they wished. None of them chose to withdraw. They were asked to sit on every other chair in order to promote examination conditions and avoid peeking. They had 1 hour to complete the test.

3.2.4 Data Analysis in the Pilot Work of the Reading Comprehension Test

To be able to investigate the face validity, reliability, item difficulty and item discrimination, three methods of data analysis were used. These methods were mean, Cronbach's alpha and item analysis.

3.2.4.1 Mean

The investigation of face validity in the pilot work of reading comprehension test employed the questionnaire with 3-point scales. The questionnaire aimed to investigate the extent that each item appeared to test the objective of the test which is the ability to scan and paraphrase specific information. Each of the three points was +1 (agree that the item appeared to test the objective), 0 (unsure that the item appeared to test the objective) and -1 (disagree that the item appeared to test the objective). The data obtained from this questionnaire were analyzed by averaging the responses from all of the five lecturers in order to investigate face validity. The possible mean could range from -1 to 1.

If the mean was 0.5 to 1.0, it meant that the lecturers agreed that the item appeared to test reading comprehension in terms of scanning and paraphrasing. On the other hand, if the mean was below 0.5, it meant that the item did not appear to test the objective and did not have face validity (Pinyoanantapong, 1984).

3.2.4.2 Cronbach's Alpha in the Pilot Work of the Reading Comprehension Test

Cronbach's alpha (α) was employed with the assistance of the SPSS 16.0 in order to investigate reliability of the reading comprehension test. It is a measure of the degree to which the items are homogenous or consistent with each other (Richards and Schmidt 2002). It is the most common way of estimating test reliability (Wiersma and Jurs, 2005). In order to use Cronbach's alpha to investigate or estimate reliability, numbers 1 to 4, representing the options a, b, c and d respectively, were typed or entered in the SPSS window of data view. In order to analyze the data, the reliability analysis command in SPSS is Analyze, Scale and then Reliability. The result of Cronbach's alpha was reported in terms of a coefficient between 0 and 1. The test was said to be reliable if the Cronbach's alpha coefficient was higher than the acceptable criterion of 0.70 (Fraenkel and Wallen, 1993).

3.2.4.3 Item Analysis in the Pilot Work of the Reading Comprehension Test

Item analysis refers to the use of various statistical procedures for analysing and improving the quality of items (Bachman and Palmer, 1996). It is the process of looking at the item-by-item responses of a test (School Improvement in Maryland, www, 1997-2009). According to classical test theory, item difficulty and item discrimination were used as the quantitative criteria for judging whether an item was good (Castillo, 1990). The item difficulty is used to investigate whether the difficulty of an item is suitable for the level of participants' knowledge or proficiency and item discrimination indicates whether the scores on the item differentiate among the abilities of the participants (MEC Scanning Office, www, 1997).

The 90 participants' answers on the reading comprehension test in this pilot work were used for the item analysis in order to check whether each item is appropriate in terms of item difficulty and item discrimination. Item difficulty involves the difficulty of each of the items. It is determined by the proportion of students correctly responding to an item. Item discrimination is the correlation between a test item and the total score (Schmidt and Embretson, 2003). It investigates the extent that a test item discriminates good and poor students. The item difficulty and item discrimination of each of the 24 items in the reading comprehension test were analyzed by Simple Items Analysis or SIA. SIA is computer software developed by Chayut Piromsombut (www, 2002).

For this pilot work of the reading comprehension test, any test items with the value of 0.20-0.80 for the item difficulty and 0.20-1.00 for the item discrimination were considered appropriate or acceptable and no change or improvement was needed (Garrett, 1966). Any items which did not meet the acceptable criteria of both item difficulty and item discrimination needed to be discarded or improved.

3.3 Phase 3: Development of Vocabulary-based Readability Index

This phase aimed to devise the vocabulary-based readability index. According to Chapter 1 and Chapter 2, there were three components of the index. These components were: 1) indication of text difficulty by LFP; 2) prediction of reading comprehension at different vocabulary size and 3) prediction of reading comprehension at different TSVK. The research methodology of this phase is presented below in terms of participants, variables, instruments, data collection procedures and data analysis.

3.3.1 Participants for the Development of Vocabulary-based Readability Index

There was one group of participants taking part in the phase. They were 102 Thai university students at Suranaree University of Technology. While they were the participants for this phase of the study, they were studying in the first class of English V course which is the last compulsory fundamental English course at the university. They were named Participant 1 to Participant 102. They had just learned and practiced reading skill in English III and English IV courses through World Class Readings 2 (Rogers, 2005) which is a reading practice book. They were about 20-23 years old. They shared similar educational backgrounds, studying English as a foreign language, having approximately 15-18 years of exposure to English and majoring in science-oriented fields. They had a wide range of levels of English language proficiency from beginners to pre-intermediate.

When the data were collected, there were 120 participants. After analyzing the data, 18 students were discarded from the study because they did not circle any word in the reading passages in the reading comprehension test or self-report on unknown words for testing TSVK but they could not translate some words in the translation test for testing TSVK. Hence, there were 102 participants in this phase.

3.3.2 Variables

As mentioned in Chapter 1, the development of the vocabulary-based readability index would involve the text-based variable which is LFP and the reader-based variables which are vocabulary size and TSVK. Additionally, according to Chapter 2, the vocabulary-based readability index was comprised of three components which are 1) indication of text difficulty by LFP; 2) prediction of reading comprehension at different vocabulary size and 3) prediction of reading comprehension at different TSVK. Therefore, the variables involved in this study were LFP, vocabulary size, TSVK and reading comprehension. These variables are presented below in terms of independent variables used to predict what would happen to a dependent variable to which it is related in some ways (Seliger and Shahomy, 1989) and a dependent variable, which is a variable that changes or is influenced according to changes in one or more independent variables (Richards and Schmidt, 2002).

3.3.2.1 Independent Variables

There are three main independent variables: 1) LFP; 2) vocabulary size; and 3) text-specific vocabulary knowledge (TSVK). Each of these variables is presented in detail below.

3.3.2.1.1 Lexical Frequency Profile

Lexical frequency profile or LFP refers to the proportions of low and high frequency words in a text. It was analyzed by RANGE – computer software (see 3.3.31). Four frequency bands: 1) the first 1000 most frequent word families; 2) the second 1000 most frequent word families; 3) the third 1000 most frequent word families; and 4) the fourth 1000 most frequent word families were used in this study.

3.3.2.1.2 Vocabulary Size

There have been several attempts to measure receptive vocabulary size and productive vocabulary size because "knowing students' receptive vocabulary size provides teachers with a gauge as to whether those students will be able to comprehend a text or a listening task, whereas knowing their productive vocabulary size provides some indication as to the degree to which students will be able to speak or write (Webb, 2008). As this study focuses on reading comprehension (see 2.2) and readability (2.3) and as mentioned in 2.5 that receptive vocabulary is the main focus, only receptive vocabulary size was tested.

Vocabulary size in this study refers to students' "quantity of vocabulary knowledge" (Gerrits, 2009) in terms of recognition of written form of words and meanings on four word lists based on the first, the second, the third and the fourth 1000 most frequent word families from the British National Corpus (BNC). The participants' vocabulary size was tested in this study by a yes/no test.

3.3.2.1.3 Text-specific Vocabulary Knowledge

Text-specific vocabulary knowledge or TSVK refers to students' knowledge on vocabulary in a text. The TSVK was one of the three independent variables investigated in the study. It was tested by two TSVK tests, which are a self-report on unknown words requiring the participants to read a passage and circle unknown words and a 32-item translation test requiring the participants to translate the underlined target words selected from the passage (see 3.3.3.3). The participants' TSVK was presented in terms of percentage of running words or tokens in a text known by individual students.

3.3.2.2 Dependent Variable

The only dependent variable in this study is reading comprehension. It was tested by a reading comprehension test presented in 3.3.3.4.

3.3.3 Instruments for the Development of Vocabulary-based Readability Index

According to Chapter 2, there were three components of vocabulary-based readability index: 1) the indication of text difficulty by LFP; 2) the prediction of reading comprehension at different vocabulary size and 3) the prediction of reading comprehension at different TSVK. In order to devise all of the three components of the index, five research instruments were used. These instruments were RANGE, a yes/no test, two tests of TSVK which are a self-report on unknown words, a translation test and a reading comprehension test. The details of each instrument are presented below.

3.3.3.1 RANGE in the Development of Vocabulary-based Readability Index

RANGE is computer software developed by Nation and Heatley (computer software, 2002) of the Victoria University of Wellington. It is the new name for the LFP analyser program. It is freely downloadable from Nation’s website at http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx. It was used in this study to analyze LFP by comparing an electronic text against vocabulary lists called baseword lists to see what words in the text are and are not in the lists and to see what percentage of the words in the text is covered by the lists (Nation, www, 2005). The baseword lists contain word families. The headword ACCOUNT is, for example, grouped with its family members which are accounted, accounting, and accounts from the list of the first 1000 most frequent word families so these three family members are counted as the same word which is ACCOUNT. These lists were saved in notepad files and entitled basewrd1.txt, basewrd2.txt, basewrd3.txt, and so on. The program categorizes vocabulary in a text(s) based on the lists that we set. For instance, we would like to check how many words in a text match the first, the second, the third, and the fourth 1000 most frequent word families from BNC. The program will categorize words in the text into five categories: four of them are categories of word families from each list and off-list - a category of words that are not in the four lists (see Table 3.6 below). “The program has self-checking routines to ensure that a word form does not occur in more than one of the baseword lists” (Wan-a-rom, 2008). This program has been prominently used in several studies such as Hirsh and Nation (1992); Laufer and Nation (1995); Coxhead (2000); Chung and Nation (2003); Nation (2006) and Wan-a-rom (2008).

The word lists used in this study are from the spoken section of BNC which is “one of the largest and most representative corpora of a single variety of English currently available” (Kennedy, 2003). Hence, the term “BNC” used in this study refers to British National Corpus from spoken section. The spoken section includes a large amount of unscripted informal conversation, recorded by volunteers selected from different age, region and social classes in a demographically balanced way, together with spoken language collected in all kinds of different contexts, ranging from formal business or government meetings to radio shows and phone-ins” (University of Oxford, www, 2005).

"High frequency words were extracted from the BNC by Leech, Rayson, and Wilson (2001) and grouped into word families by Nation" (Cobb, 2007). There are fourteen 1000 word-family lists revised from Nation (2006). They were the only BNC word lists available when the LFP of reading passages were analyzed and when the yes/no test used in this study was constructed.

The lists in Nation (2006) were based on the whole 100,000,000 token in BNC. However, Nation and Beglar (2007) states that the largely formal written nature of the BNC strongly affected the high frequency levels. According to Nation, the tokens like cat, hello, sun, worse which frequently occur in everyday life are gathered in the 4th 1000 rather than at a higher frequency level and some very formal words like civil and commission appear in the first 1000 word list from Nation (2006). Therefore, Nation and Beglar decided to revise the first twelve 1000 word lists using word family range and frequency figure from only the 10 million tokens from the spoken section of the BNC (Nation, 2006; Nation and Beglar, 2007). Nation and Beglar (2007) states that this revision resulted in a more sensible ordering although the changes were not large. Although it may seem a little strange to use a spoken corpus-based ordering for a test of written receptive knowledge, it was felt that the spoken ordering more closely represented the order in which the intended students might learn the words (Nation and Beglar, 2007).

The word lists are freely downloadable from http:// www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx and http://lextutor.ca/vp/bnc/nati on14/. The former webpage provides these word lists that come with the RANGE program. The latter one presents the word lists as alternative word lists for language teachers and students who want to analyze LFP or to break texts down by word frequencies based on BNC (Cobb, www, 2006). These word lists are used in Cobb (2007) in order to investigate the vocabulary demands of L2 reading by analyzing LFP and in Nation and Beglar (2007) to construct a receptive vocabulary size test.

These word lists are categorized and counted based on word families because learners beyond a minimal proficiency level have some control of word building devices and are able to see that there is both a formal and a meaning relationship between regularly affixed members of a word family" (Nation and Beglar, 2007). There is also some evidence that the word family is a psychologically real unit (Nagy, Anderson, Schommer, Scott and Stallman, 1989; Bertram, Baayen and Schreuder, 2000; Bertram, Laine and Virkkala, 2000).

The lists are made to run in the RANGE program. In order to analyze LFP by RANGE, we need to follow the steps below.

1. Go to Paul Nation's website at http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspxwe in order to download the program. We need to do it for the first time only.
2. Save the text or texts that you want to run the program on text files which are files with .txt in the name of the files.
3. Double click on the RANGE icon in Windows explorer to access to the program.
4. Open the File menu in RANGE and choose the heading Open.
5. Select the file or files you want to run the program over. These files must all be text files. In this study, the analyzed text files were

reading passages scanned from reading practice books available in Thailand by using Optical Character Recognition or OCR software.

6. After you have selected the files, click on Open, go to the File menu again and choose Save. Type the name of the file that you want to save the results to, and click on Open.
7. Look at the list of options at the bottom of the RANGE window. You can change these options or leave the options as they are. In this study, the number of Baseword Files is set at 4 in order to relate it to the yes/no test which consists of words from these four Baseword Files. The first four baseword files or word lists are the first, the second, the third and the fourth 1000 most frequent word families categorized from BNC. Click the button Process Files which is below the file list in the RANGE window.
8. Look at the results file using a word processor like MS-Word or notepad. The results file will be the name you chose plus _range.txt, for example results_range.txt. The results from the analysis of LFP of Passage One are given as follows.

Table 3.6 Lexical Frequency Profile of Passage One in the Reading Comprehension Test with Four Passages

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
one	139/95.86	73/92.41	68
two	6/4.14	6/7.59	6
three	0/0.00	0/0.00	0
four	0/0.00	0/0.00	0
Off-list	0/0.00	0/0.00	?????
Total	145	79	74

The program presents the numbers and percentages of words in terms of tokens and types and word families in a target English text coming from each of the word lists and those which are not recognised in the lists or off-list words. A token or running word is “any occurrence of a word form in the text, regardless of whether it is occurring for the first or the n^{th} time. A type is any word form which is counted only once, regardless of how many more times it might occur. A word family consists of a base word and all its derived and inflected forms (Bauer and Nation, 1993). Since the analysis of LFP in this study aims to show the percentage of running words in each passage covered by each word-family list, the word “token” is used as a counting unit in this study.

From Table 3.6, we can see that both numbers and percentage of occurrences are given. The number before the slash (/) refers to numbers of occurrences in each word list and the number after the slash (/) refers to the percentage of occurrences. The above figure shows that 139 tokens out of 145 tokens or running words in Passage One are in word list one or the first frequency band consisting of the first 1000 most frequent word families and these 139 words make up 95.86% of the total tokens in the text.

3.3.3.2 Yes/no Test in the Development of Vocabulary-based Readability Index

The yes/no test used in this phase was piloted in Phase 1 which is the pilot work of the yes/no test. There were four levels in the test (Level 1, Level 2, Level 3 and Level 4). Each of the four levels consisted of 30 target words randomly selected from the first, the second, the third and the fourth 1000 most frequent word families respectively. Each level was also comprised of 20 pseudowords constructed by Meara and his colleagues at University of Wales Swansea (Cobb, www, 2006). The yes/no test used in the study is shown in Appendix A. Like the yes/no test in Phase 1, there were two forms of the yes/no test (Test A and Test B) differing in order so as to avoid cheating.

The yes/no test was scoring by $h - f$ which is the scoring method used in the pilot work of the yes/no test. To be able to calculate $h - f$, the individual's selected target words and their selected pseudowords were counted. The number of selected target words was divided by the total number of target words (30) in each level in order to calculate the hit alarm rate (h). The number of selected pseudowords was divided by the total number of pseudowords (20) in each level in the yes/no test so as to calculate the false alarm rate (f). Then the false alarm rate was subtracted from the hit alarm rate so as to get the yes/no test score. The scores on each level in the yes/no test could range from -1 to 1.

3.3.3.3 Tests of TSVK in the Development of Vocabulary-based Readability Index

TSVK was tested by two research instruments. These instruments were a self-report on unknown words and a translation test. This section aims to present the details of these two instruments as well as how to calculate TSVK.

The self-report on unknown words was used to test the 102 university students' unknown words in individual reading passages. These students taking part in this phase as participants were asked to read each of the four reading passages in the reading comprehension test and circle unknown words (see Appendix D). The circled words were counted and compared with the answers on the translation test.

The translation test was constructed by asking 49 undergraduates in Faculty of Education at Silpakorn University to circle unknown words in the reading passages from the reading comprehension test. Eight words which were unknown in each passage by most of the undergraduates were included in this translation test The selected words were: ourselves, certain, amount, situation, decide, marriage, brain, choosing, jeopardy, draught, sneezes, proximity, influences, exposed, complain/complains, coughs, establish/establishing, crayons, kindergarten, maturity, arouse, routine, schedule, generate, semiconductors, purchases, deficit, machinery, revealing, petroleum, decline, and steadily. In the translation test, these selected words were presented in contexts copied from the reading passages (see Appendix F).

The individual participants' reported unknown words in the self-report on unknown words were counted. A comparison was then made between each individual's circled words and their answers in the translation test. There were five kinds of comparison as follows:

1) A word was correctly translated and not circled in the reading passages as unknown. That is to say, the participant considered the word as known and translated the words in the translation test correctly.

2) A word was left not translated at all and circled in the text. In this case, the participant did not know the meaning of the word.

3) Some words were circled or reported as unknown. However, one or more words were mistranslated or not translated but not underlined in the text. This shows that the participant reported the word as known but he/she did not actually know the meaning of some of the words.

4) A word was circled but it was correctly translated in the translation test. It means that the participants knew the word but he/she underestimated his/her vocabulary knowledge so he/she reported the word as unknown although he/she knew the meaning of the word.

5) The participant did not report any unknown words but he/she could not translate some or all of the words in the translation test. It reveals that the participant did not know the meaning of the words but he/she overestimated his/her vocabulary knowledge so he/she did not report the words as unknown.

From the comparisons above, we can see that some participants seemed to be aware of their vocabulary knowledge while others might not. Some of them tended to be so confident that they overestimated their vocabulary knowledge while others did not have much confidence so that they underestimated theirs. This caused some discrepancies between the results of the two tests.

For the first two kinds of comparisons, the words reported as unknown were counted and calculated into percentage. For the third comparison, the calculation employed by Laufer (1989) was used. This calculation is expressed in two main stages:

Stage 1: The number of words known in a text is:

Total tokens in a text – [tokens reported as unknown + ((tokens reported as unknown × (number of discrepancies × 100/40))/100)]

After the participants completed the self-report on unknown words and the translation test, the individual's reported unknown words were compared with their answers in the translation test in order to check whether there were any tokens that were not reported as unknown but could not be translated in the translation test. A token or running word that was not reported as unknown or a token that was not circled in the self-report on unknown words for testing TSVK but not translated or mistranslated in the translation test for testing TSVK is called a discrepancy. In the first stage of the calculation, the number of discrepancies was multiplied by 100 and divided by 40 in order to calculate percentage of tokens that were not reported as unknown. When multiplied by the number of reported unknown tokens and divided by 100, we got the number of tokens that was not reported as unknown but should have been. Hence, we add the number of reported unknown tokens to those that should have been reported so as to calculate how many words were really unknown. This number was subtracted from the number of tokens in the reading passage in order to see how many tokens were known.

According to Laufer's (1989) first stage of calculating TSVK, Laufer was a little vague about the number of discrepancies × 100/40 because how she got number 40 was not known. She might get this number from the number of tokens in texts in her study or number of words in the translation test. It is questionable whether the calculation used in her study was applicable to the present study or not because we did not strictly follow Laufer's procedures, namely number of words in the translation test. Laufer asked a small group of students at equivalent level as the participants in her study asking for a larger number of words (except the very frequent ones and the function words) and then selected the ones that were not familiar (B. Laufer, personal communication, May 9, 2007) while only 8 words which were reported unknown by most of the 49 undergraduates at Silpakorn University were used in the present study. Hence, it is uncertain that the calculation accurately calculated TSVK for the present study. Furthermore, it seems ridiculous when the calculation requires us to multiply by 100 and then divide by 100. However, it is the only calculation of TSVK so we will carry on with it.

Stage 2: The text-specific vocabulary knowledge (TSVK) is:

The number of tokens known in the text × 100/Total number of tokens in the text

In the second stage, the number of tokens known in the reading passage or the result from the first stage was multiplied by 100 and divided by total number of tokens in the reading passage in order to convert the number of known tokens into percentage of the total number of tokens in the text. This percentage is called TSVK.

The above calculation is also used for the fourth kind of comparison when a participant underestimates his/her vocabulary knowledge. However, it needs some adaptation. From the above calculation, the number of discrepancies is positive but when a participant underestimates his/her vocabulary knowledge, this number needs to be negative.

For the fifth comparison, the number of words reported as unknown is zero. When we use the above calculation to calculate the TSVK, the TSVK will be 100. It means that the participant knows every word in the text. It is acceptable when the participant can translate every word in the translation test. If he/she cannot, it can be concluded that he/she overestimates their vocabulary knowledge. Therefore, he/she will be discarded from the study.

3.3.3.4 Reading Comprehension Test in the Development of Vocabulary-based Readability Index

The reading comprehension test used in this phase for the development of the vocabulary-based readability index had been piloted in the pilot work of the reading comprehension test (see Phase 2) in order to investigate reliability, face validity, item difficulty and item discrimination. According to Phase 2, the test was constructed based on the test specifications presented in Table 3.5. It aimed to test participants' reading comprehension or the ability to scan and paraphrase specific information (see 2.2). The test consisted of four reading passages and six test items or questions (see Appendix D). Each of the four reading passages had different LFP. Passage One had more than 95% of tokens from the first 1000 most frequent word families. Passage Two consisted of more than 95% of tokens from the first and the second 1000 most frequent word families and so on. The information about LFP including number of tokens and percentage of tokens from each BNC word list was presented in Table 3.3 (see above).

To be able to score the reading comprehension test, each of the four sections in the reading comprehension test was scored separately. Each item was worth one mark. The participant whose answer matched the answer key would get one mark and the participant who got a wrong answer would get zero. Since there were six items in each section in the reading comprehension test, the possible scores on the test ranged from 0 to 6.

3.3.4 Data Collection Procedures in the Development of Vocabulary-based Readability Index

Before the data were collected, the 102 participants, who were taking the fifth fundamental English course at Suranaree University of Technology, were informed the purpose of the data collection that is to devise the vocabulary-based readability index. They were also given the opportunity to not participate or to withdraw at any time they wished. All of them agreed to take part in the study. They were asked to sit on every other chair in order to promote examination conditions and avoid peeking.

The participants were asked to complete the yes/no test, the self-report on unknown words for testing TSVK, the reading comprehension test and the translation test for testing TSVK. They had 90 minutes to complete the tests. They were not allowed to use any dictionaries.

3.3.5 Data Analysis in the Development of Vocabulary-based Readability Index

This part presents the methods of data analysis employed in the development of vocabulary-based readability index. The main data analysis carried out here was: descriptive statistics including mean and standard deviation, three-way ANOVA and Scheffé Test, Pearson product-moment correlation or Pearson *r* and simple regression analysis. These statistical methods were calculated with the assistance of the

Statistical Package for the Social Science (SPSS) version 16.0 software. Each method is presented in detail below.

3.3.5.1 Descriptive Statistics

Descriptive statistics including mean and standard deviation was employed for an overall picture of the participants' reading comprehension scores. The mean is a measure of central tendency which refers to the middle value or a typical value of the data (Internet Center for Management and Business Administration, www, 1999-2010). It is the sum of all the scores divided by the total number of participants. Standard deviation is a measure of dispersion or how far the scores are from the mean score. It is the square root of variance. With the assistance of SPSS 16.0, the command for calculating the mean and the standard deviation was Analyze, Descriptive Statistics and then Descritives.

3.3.5.2 Three-way ANOVA and Scheffé Test

Three-way Univariate Analysis of Variance (ANOVA) in General Linear Model in SPSS was used to test seven null hypotheses:

1) There is no difference of the mean reading scores when the LFP is varied;
2) There is no difference of the mean reading scores when the TSVK is different;
3) There is no difference of the mean reading scores when the vocabulary sizes are varied;
4) There is no interaction between LFP and TSVK (LFP*TSVK);

5) There is no interaction between LFP and vocabulary sizes (LFP*vocabulary sizes);

6) There is no interaction between TSVK and vocabulary sizes (TSVK*vocabulary sizes); and

7) There is no interaction between LFP, TSVK and vocabulary sizes (LFP*TSVK*vocabulary size).

In brief, the first three hypotheses concern the comparisons of the actual reading comprehension scores (dependent variable) across the reader-based variables (independent variables) which are LFP, vocabulary size and TSVK. The other four hypotheses have something to do with the effect of interaction between the independent variables (LFP, TSVK and vocabulary sizes) on the dependent variable (reading comprehension scores).

Since the three-way ANOVA only reveals whether there is a significant difference of mean scores when the independent variables are varied and whether interactions of the independent variables have any effect on the dependent variable, it could not pinpoint the differences among groups. If there is a significant difference found in the result of the three-way ANOVA, post-hoc Scheffé tests would be used to indicate which pairs of the groups under the variables contribute to the overall differences. In order to use three-way ANOVA to analyze the data, the command in SPSS is Analyze, General linear model and then Univariate.

3.3.5.3 Pearson Product-Moment Correlation in the Development of Vocabulary-based Readability Index

Pearson product-moment correlation or Pearson *r* is a measure of correlation between two continuous variables (Richards and Schmidt, 2002). It was used in the development of vocabulary-based readability index as follows: 1) to investigate the correlations between the three ways of indicating text difficulty by LFP and the results of Flesch readability formula and Flesch-Kincaid readability test and 2) to investigate the correlations among the results of the three ways of using LFP to indicate text difficulty. These correlations were investigated so as to select the most suitable method of using LFP to indicate text difficulty.

3.3.5.4 Simple Regression Analysis

Simple regression analysis is a statistical technique for estimating or predicting a value for a dependent variable from an independent variable (Richards and Schmidt, 2002). As mentioned in 3.3.2, the dependent variable was reading comprehension and the independent variables were vocabulary size and TSVK. Simple regression analysis was used in this study to form equations for predicting or estimating reading comprehension scores at different vocabulary size and predicting reading comprehension scores at different TSVK. With the assistance of SPSS 16.0, the simple regression was calculated by going to Analyze, Regression and then Linear.

3.4 Phase 4: Validation of Vocabulary-based Readability Index

This phase aimed to validate the three components of the vocabulary-based readability index in order to answer the research question "To what extent would the purely vocabulary-based readability index be valid?". The three components were: 1) the indication of text difficulty by LFP; 2) the prediction of reading comprehension at different vocabulary size and 3) the prediction of reading comprehension at different

TSVK. This section includes the details of research methodology in terms of participants, instruments, data collection procedures and data analysis.

3.4.1 Participants for the Validation of Vocabulary-based Readability Index

According to Table 3.1, four groups of participants were asked to take part in this phase of the study in order to serve three main purposes: 1) three Thai students and three Thai lecturers were asked to take part in the investigation of face validity of the three components of the index; 2) forty-five university students and five lecturers were asked to take part in the investigation of concurrent validity of LFP as an indicator of text difficulty and 3) eighty and thirty university students were asked to take part in the investigation of predictive validity of the predictions of reading comprehension at different vocabulary size and at different TSVK. The details of the four groups of participants are described below.

Six participants were asked to take part as participants in the investigation of face validity of the three components of the index. Three of them were Thai university students and the other three were Thai lecturers. The students were third year students at a public university in Thailand. They had completed four fundamental English courses. They had also experienced text selection from the Internet and the library. They were named S1, S2 and S3 in the study. For the lecturers, they had at least three years of teaching experience. All of them had selected reading passages for their students in fundamental English courses. They had exposed to English language for about 15-18 years. They were named L1, L2 and L3.

Forty-five university students at Kasetsart University took part in the investigation of concurrent validity. They were studying the fifth fundamental English course focusing on reading when the data were collected. They had a three-hour class per week. They were trained the reading skill by an in-house material. They had studied four fundamental English courses at the university. They had about 15-18 years of exposure to English. They were named Student 1-Student 45. Five lecturers were also asked to take part in the investigation of concurrent validity of the vocabulary-based readability index. They had been responsible for English language teaching at their university for at least two years. All of them had experienced selecting texts for their students. They were called Teacher 1-Teacher 5 in the study.

Eighty university students were asked to take part in this phase in order to investigate predictive validity. They were studying English III course which is the third fundamental English course at Suranaree University of Technology. This course focuses on reading. The textbook used in this course is World Class Readings 2 (Rogers, 2005). They had about 14-17 years of exposure to English. They were given identification codes ST1-ST80.

Thirty students studying the third fundamental English course at Silpakorn Univeristy were also asked to take part in the investigation of predictive validity. They were studying the third fundament English course at the university. They had about 14-17 years of exposure to English. They were named Student A-Student AD.

3.4.2 Instruments for the Validation of Vocabulary-based Readability Index

In order to be able to answer the research question mentioned in 1.3 (To what extent would the purely vocabulary-based readability index be valid?), six main research instruments were used. These instruments were: 1) RANGE; 2) a yes/no test; 3) two self-reports on unknown words; 4) two translation tests; 5) two reading comprehension tests and 6) three questionnaires. Each instrument is presented below.

3.4.2.1 RANGE in the Validation of Vocabulary-based Readability Index

RANGE is computer software. It compares an electronic text against vocabulary lists in order to see what words in the text are and are not in the lists and to see what percentage of the words in the text is covered by the lists (Nation, www, 2005). It was used in the study to analyze LFP based on four word lists from BNC spoken section. (see 3.3.3.1).

3.4.2.2 Self-reports on Unknown Words for Testing TSVK in the Validation of Vocabulary-based Readability Index

Two self-reports on unknown words were used in this phase of the study. Both of them required the participants to read the reading passages and circle unknown words. One of them was used to test TSVK in the reading passages in the reading comprehension test with five passages from TOEIC tests (see Appendix G) and the other one was employed to test TSVK in the reading passages in the reading comprehension test with four passages (see Appendix D).

3.4.2.3 Translation Tests for Testing TSVK in the Validation of the Vocabulary-based Readability Index

Two translation tests were also used to test TSVK. One of them was employed to test TSVK in the reading passages in the reading comprehension test with five passages from TOEIC tests and the other one was used to test TSVK in the reading passages in the reading comprehension test with four reading passages. Each of them was constructed by asking undergraduates to read the reading passages in each reading comprehension test and circle words in the passages that they did not know well enough to say what they mean. Eight words that were unknown in each passage by most participants were included in the translation test (B. Laufer, personal communication, May 9, 2007). In the translation tests, the selected words were presented in contexts copied from the reading passages.

Since there were five reading passages in the reading comprehension test from TOEIC tests and eight words in the translation test were selected from each reading passage, there were 40 words in the translation test for testing TSVK for this reading comprehension test. These words were tunnel, relieve, preliminary, tremendous, struggling, opponents, temporary, shortage, classy, viable, sacks, disposables, durable, carryall, catchy, commissioned, attractions, inaugurated, generated, dispense, construction, complex, recreational, simulated, amateurs, comprehensive, exposures, sitter, integrate, props, illustrate, refreshments, gradual, recovery, protracted, recession, substantial, mortgage, boost and declining (see Appendix H).

As there were four reading passages in the other reading comprehension test and eight words were selected from each reading passage, there were 32 words in the translation test for testing TSVK in the reading comprehension test with four passages. The selected words were: ourselves, certain, amount, situation, decide, marriage, brain, choosing, jeopardy, draught, sneezes, proximity, influences, exposed, complain/complains, coughs, establish/establishing, crayons, kindergarten, maturity, arouse, routine, schedule, generate, semiconductors, purchases, deficit, machinery, revealing, petroleum, decline, and steadily (see Appendix F).

According to 3.3.3.3, the reported unknown words were compared with the answers in the translation tests for testing TSVK in order to get TSVK. If there were no discrepancies or tokens that were not reported as unknown in the self-report on unknown words but could not be accurately translated in the translation test, the reported unknown words that were not correctly translated in the translation test were counted. This number of reported unknown words was subtracted from the total number of tokens in the passage. Then the result was converted to percentage in order to get TSVK. If there were some discrepancies, Laufer's (1989) two stages of calculation presented in 3.3.3.3 should be used. By way of illustration, let us look at how to get the TSVK step by step. Suppose that a participant reported 20 unknown tokens and there were 10 discrepancies. When the word count available in Microsoft Word 2003 was employed, the total number of words was 150.

1. Compare each participant's reported unknown words with his/her answers in the translation test. From the comparison, we find 10 discrepancies i.e. ten words are said to be known but the participant does not exactly know the words so they cannot translate them in the translation test.

2. The calculation starts with calculating the percentage of discrepancies which is (number of discrepancies × 100/40) or 10 ×

100/40 = 25%. This shows what percentage of vocabulary is unknown but not reported.

3. We need to multiply the result that we got from Stage 2 (25%) by the number of reported unknown words. We will get the number of words that was not reported as unknown but should have been. However, the number that we get from the last stage is in terms of percentage so it is impossible for us to multiply a percentage with a number. We need to divide the percentage by 100 in order to make a percentage a number so the calculation is 20 × 25/100 = 5. From the calculation, 5 is the number of words that should have been reported as unknown but they participant did not.

4. We, then, need to add 5 (the number that we got from Stage 3) to the number of words reported as unknown so as to see how many words which are really unknown to the participant. This means the number of words that he/she should have reported as unknown which is 20 + 5 = 25.

5. In order to see the number of the participant's known words, we need to subtract the result from the calculation in Stage 4 (25) from the number of total words in a text. As mentioned earlier, there are 150 words so the number of words known is 150 – 25 = 125.

6. The number of known words in Step 5 (125) is calculated by percentage which is 125 × 100/150 = 83.33%. It means that the participant knows 83.33% of tokens in the text.

3.4.2.4 Reading Comprehension Tests in the Validation of the Vocabulary-based Readability Index

Two reading comprehension tests were used in the present study. Both of the reading comprehension tests were used to test participants' reading comprehension or their ability to scan and paraphrase specific information (see 2.2). One of the test consisted of five reading passages and 20 items (see Appendix G). The first four reading passages and sixteen questions were taken from Tactics for TOEIC: Listening and Reading Tests Practice Test 1 and Test 2 (Educational Testing Service, 2007a, 2007b). These practice tests are authorized by Educational Testing Service (ETS) – an organization organizing TOEIC tests all over the world. The practice tests contain official TOEIC test items. The last reading passage and the last four items were taken from a TOEIC practice book (Prachyawisan, 2005). This reading comprehension test is called the reading comprehension test with five passages from TOEIC tests throughout the study. TOEIC was used in the present study because some of the texts in TOEIC tests contain 95% of tokens from the first four frequency bands while all of the TOEFL and IELTS texts seem to be much more difficult because 95% of tokens from these texts are from the first five or more frequency bands. The LFP of the five reading passages in the test is shown below.

Table 3.7: Lexical Frequency Profile of Five Reading Passages from TOEIC Tests

	Passage One	Passage Two	Passage Three	Passage Four	Passage Five
Tokens/% in the first BNC word list	234/81.82	234/76.72	176/78.22	218/78.14	90/78.26
Tokens/% in the second BNC word list	25/8.74	40/13.11	33/14.67	33/11.83	15/13.04
Tokens/% in the third BNC word list	12/4.20	15/4.92	2/0.89	6/2.15	3/2.61
Tokens/% in the fourth BNC word list	2/0.70	2/0.66	3/1.33	15/5.38	3/2.61
Off-list	13/4.55	14/4.59	11/4.89	7/2.51	4/3.48

The numbers and the percentage of tokens from each of the five BNC word lists are given in Table 3.7. The number before the slash (/) refers to numbers of tokens from each BNC word list and the number after the slash (/) refers to the percentage of tokens from each BNC word list. By way of illustration, let us look at LFP in Passage One presented in the above table. It is shown that 234 tokens out of 286 tokens or running words in Passage One are from the first BNC word list consisting of the first 1000 most frequent word families and these 234 tokens make up 81.82% of the total tokens in Passage One. We can see from Table 3.7 that all of the five passages contained 95% or more from the first four BNC word lists (95.45%, 95.41%, 95.11%, 97.49% and 96.52% respectively).

The other reading comprehension test consisted of four reading passages and 24 items (see Appendix D). This test is named the reading comprehension test with four passages. The test was piloted in the pilot work of reading comprehension test and used in the development of the vocabulary-based readability index. The reading passages were selected from reading practice books, namely, Bussayasiri (1989), Prachyawisan (2005) and Bell (2006). They were selected because each of them consisted of 95% of tokens from the first, the first and the second, the first, the second and the third or the first, the second, the third and the fourth 1000 most frequent word families. The information about LFP including number of tokens and percentage of tokens from each BNC word list was presented in Table 3.3 (see above).

3.4.2.5 Questionnaires in the Validation of the Vocabulary-based Readability Index

Three questionnaires were employed in the validation of the vocabulary-based readability index. They were used to serve three different purposes: 1) to investigate face validity of the three components of the index; 2) to investigate the concurrent validity of LFP as an indicator of text difficulty and 3) to investigate predictive validity of predictions of reading comprehension by vocabulary size and TSVK. Each of them is described in detail below.

3.4.2.5.1 Questionnaire for Investigating Face Validity of Vocabulary-based Readability Index

According to Chapter 2, there were three components of the vocabulary-based readability index: 1) indication of text difficulty by LFP; 2) prediction of reading comprehension at different vocabulary size and 3) prediction of reading comprehension at different TSVK. In order to investigate face validity of these three components, a questionnaire was used to survey the three university students' and the three lecturers' opinions whether they agreed that the LFP looked good as an indicator of text difficulty and vocabulary size and TSVK looked good as predictors of reading comprehension. The questionnaire consisted of three items and each item was a 5-point scale. The scale was valued 1 (extremely disagree); 2 (disagree); 3 (unsure); 4 (agree) and 5 (extremely agree). The questionnaire was written in Thai. The Thai and English versions of the questionnaire are shown in Appendix I.

3.4.2.5.2 Questionnaire for Investigating Concurrent Validity of LFP as an Indicator of Text Difficulty

The questionnaire for ranking text difficulty was used to investigate concurrent validity of LFP. It was used to elicit 45 university students' and 5 lecturers' rankings of the four reading passages in the reading comprehension test in terms of text difficulty. There was only one item in this questionnaire. The questionnaire was written in Thai so it was easy for the participants to read and understand what they were asked to do. The Thai and English versions of the questionnaire are given in Appendix J. The students and the lecturers were asked to rank the four reading passages in the reading comprehension test with four passages in terms of text difficulty from 1 (the easiest) to 4 (the most difficult).

3.4.2.5.3 Questionnaire for Investigating Predictive Validity of the Prediction of Reading Comprehension

The questionnaire for rating text difficulty was used to investigate predictive validity of the prediction of reading comprehension by vocabulary size and TSVK. It consisted of four 5-point rating scales. Each of the scales was valued 1 (too difficult), 2 (difficult), 3 (optimal), 4 (easy) and 5 (too easy). It requires the participants to rate the extent of difficulty of each of the reading passages in the reading comprehension test with four passages. The questionnaire was written in Thai. The Thai and English versions of the questionnaire are given in Appendix K.

3.4.3 Data Collection Procedures in the Validation of Vocabulary-based Readability Index

As mentioned in 3.4.1, there were four groups of participants in the validation of the vocabulary-based readability index. They were asked to take part in this phase in order to serve three purposes: 1) to investigate face validity of the three components of the vocabulary-based readability index; 2) to investigate concurrent validity of LFP as an indicator of text difficulty and 3) to investigate predictive validity of the prediction of reading comprehension. The data collection procedures are presented below based on each of the three purposes.

3.4.3.1 Data Collection Procedures for Investigating Face Validity of Vocabulary-based Readability Index

Before the data for investigating face validity were collected, each of the three students and the three lecturers was informed the purposes of the three components of the vocabulary-based readability index and the instructions of using each of the components. They were then asked to complete the questionnaire for investigating face validity of the vocabulary-based readability index without any time limit. They were told to choose a number from 1 (extremely disagree) to 5 (extremely agree) on the scale that represented the extent that they agreed that 1) the LFP looked good as an indicator of text difficulty; 2) vocabulary size looked good as a predictor of reading comprehension and 3) TSVK looked good as a predictor of reading comprehension.

3.4.3.2 Data Collection Procedures for Investigating Concurrent Validity of LFP as an Indicator of Text Difficulty

This data collection aimed to investigate concurrent validity of LFP as an indicator of text difficulty. Prior to the data collection, 45 students and 5 lecturers were told the purpose of the data collection. They were asked to look through each of the four reading passages in the reading comprehension test and rank these passages in terms of text difficulty. They were required to write number 1 (the easiest) to number 4 (the most difficult) in the questionnaire in order to rank the difficulty of the four reading passages. They had 40 minutes to complete the questionnaire.

3.4.3.3 Data Collection Procedures for Investigating Predictive Validity of Predictions of Reading Comprehension

In order to investigate predictive validity of reading comprehension by vocabulary size and TSVK, the data were collected from two groups of participants. There were 80 and 30 university students respectively.

Before the data were collected, the participants in both groups were informed the purpose of the data collection. They were also given the opportunity to not participate or to withdraw at any time they wished. They were asked to sit on every other chair in order to promote examination conditions and avoid peeking.

Eighty students were asked to take part in the data collection because we would like to know the extent that vocabulary size or TSVK accurately predicts reading comprehension scores on reading comprehension questions from a standardized test like TOEIC. When the four tests which were: 1) the yes/no test; 2) the self-report on unknown words for testing TSVK in the reading comprehension test with five passages from TOEIC tests; 3) the reading comprehension test with five passages from TOEIC tests and 4) the translation test for testing TSVK in the reading comprehension test with five passages from TOEIC tests were administered, the participants were told what they were supposed to do in each test. They had 90 minutes to complete all of the tests.

Thirty students took part in this phase of the study because we would like to investigate the extent that vocabulary size or TSVK accurately predicts how easy or difficult a passage is for individual students to comprehend it. After the yes/no test, the self-report on unknown words for testing TSVK in the reading comprehension test with four passages, the reading comprehension test with four passages, the translation test for testing TSVK in the reading comprehension test with four passages and the questionnaire for rating text difficulty were distributed to the students, they were told what they were required to do in each of the tests and the questionnaire for rating text difficulty. They had 100 minutes to complete the tests and the questionnaire.

3.4.4 Data Analysis in the Validation of Vocabulary-based Readability Index

This part presents the methods of data analysis employed in this study. The main data analysis carried out here are: frequency, Spearman's Rank Correlation or Spearman's rho and Pearson product-moment correlation or Pearson *r*. They were calculated with the assistance of the Statistical Package for the Social Science (SPSS) version 16.0 software.

3.4.4.1 Frequency for the Investigation of Face Validity of Vocabulary-based Readability Index

Frequency was used in this phase to show the number of the three students' and the three lecturers' responses on each item in the questionnaire for investigating face validity. The maximum was 3 and the minimum was 0.

3.4.4.2 Spearman's Rank Correlation for the Investigation of Concurrent Validity of LFP as an Indicator of Text Difficulty

Spearman's rank correlation or Spearman's rho was used to investigate the correlations of rankings of the reading passages in the reading comprehension test with four passages in terms of text difficulty. The results were used to investigate the concurrent validity of the indication of text difficulty by LFP. In order to do so, Spearman's rho was used to investigate the rank correlations between the rankings of the reading passages in the reading comprehension test with four passages ranked by LFP, the 45 university students and the 5 lecturers.

3.4.4.3 Pearson Product-Moment Correlation for the Validation of the Vocabulary-based Readability Index

Pearson *r* was used in this phase of the study to investigate the concurrent validity of the indication of text difficulty by LFP and predictive validity of the predictions of reading comprehension at different vocabulary size and at different TSVK. In order to do so, Person *r* was used to investigate the following correlations:

1) For the investigation of concurrent validity of LFP as an indicator of text difficulty, the correlations between the percentage of tokens from the first three

frequency bands which is the results of LFP and the results of two readability formulas which are Flesch readability formula and Flesch-Kincaid readability test were investigated.

2) For the investigation of the predictive validity of the prediction of reading comprehension scores at different vocabulary size and at different TSVK, the correlations between the reading comprehension scores predicted by vocabulary size and by TSVK and the reading scores of the 80 university students were investigated.

3) For the investigation of the predictive validity of the prediction of the ease or difficulty of a text or how easy or difficult a text is for individual students to comprehend the text by vocabulary size and TSVK, the correlations between the ease or difficulty predicted by vocabulary size and TSVK and the 30 participants' rating of text difficulty were investigated.

3.5 Summary

This chapter described the research methodology including participants, instruments, data collection procedures and data analysis employed in the four phases of the present study. These phases were 1) the pilot work of the yes/no test; 2) the pilot work of the reading comprehension test; 3) the development of the vocabulary-based readability index and 4) the validation of the vocabulary-based readability index. The results of Phase 1 and Phase 2 will be presented in Chapter 4 and the results of Phase 3 and Phase 4 will be shown in Chapter 5.

CHAPTER 4

THE RESULTS OF THE PILOT WORK OF THE YES/NO TEST AND THE PILOT WORK OF THE READING COMPREHENSION TEST

The main purpose of this chapter is to present the results of the pilot work of the yes/no test and the pilot work of the reading comprehension test. This chapter is organized into two sections. The first section reports the results of the pilot work of the yes/no test in terms of test reliability, concurrent validity and construct validity. The second section deals with the presentation of results of the pilot work of the reading comprehension test in terms of face validity, test reliability, item difficulty and item discrimination.

4.1 The Results of the Pilot Work of the Yes/no Test

This section presents results of the pilot work on the yes/no test. It is divided into three parts: 1) summary of scores on the yes/no test; 2) investigation of reliability of the yes/no test and 3) validation of the test in terms of concurrent validity and construct validity.

4.1.1 Summary of Scores on the Yes/no Test

This section aims to present the summary of scores on the yes/no test in order to give an overview of the participants' performance on the yes/no test and exemplify how the yes/no test scores were presented in the study. After using *h-f* which is the scoring method of the yes/no test used in this study, the calculated scores could range from -1 to 1. The 75 participants' scores on the yes/no test are presented in Appendix L. In order to present the overall picture of the participants' performance on the yes/no test, the scores are presented in terms of the mean scores from each level of the four levels in the yes/no test (Level 1, Level 2, Level 3 and Level 4), standard deviation, maximum and minimum scores.

Table 4.1: Summary of scores from the yes/no test

	Level 1	**Level 2**	**Level 3**	**Level 4**
Mean	0.64	0.25	0.18	0.10
Standard deviation	0.196	0.186	0.186	0.125
Minimum	0.00	0.00	-0.10	-0.05
Maximum	0.90	0.82	0.70	0.50

From Table 4.1, "Level 1" refers to the first 1000 most frequent word families, "Level 2" is the second 1000 most frequent word families, "Level 3" is the third 1000 most frequent word families and "Level 4" refers to the fourth 1000 most frequent word families in the BNC. The mean scores are the average of scores or the sum of scores of the participants on each of the four levels of the yes/no test (Level 1, Level 2, Level 3, Level 4), divided by the total number of participants taking the test (75).

The mean score on Level 1 consisting of target words from the first 1000 most frequent word families in the yes/no test was the highest (0.64) while the mean score on Level 4 consisting of target words from the fourth 1000 most frequent word families was the lowest (0.10). This implies that most of the participants tend to get the highest score on Level 1 and the scores gradually decreased when the words in the yes/no test are less familiar.

The standard deviation was used to measure dispersion of scores or how far the scores are from the mean score. The standard deviations (0.196, 0.186, 0.186 and 0.125) shown in Table 4.1 were close to zero. It should have been concluded that the participants' scores are not far from the mean scores. However, the standard deviations do not tell us the variation of the scores. It only tells us the variation as the relative number compared to the mean. When we look at the mean scores presented in Table 4.1 (0.64. 0.25, 0.18 and 0.10), the standard deviations were close to the mean scores. Comparing to the means, the standard deviations of scores especially scores on Level 3 and Level 4 show a lot of variations.

To make the scores on the yes/no test more practical and more plausible, the yes/no test scores from the scoring method (*h*-*f*) were multiplied by 1,000 which is the size of frequency band or number of word families in a word list in order to convert the scores into number of word families a participant knows out of 1,000 word families (W. F. Admiraal, personal communication, February 11, 2009).

4.1.2 Reliability of the Yes/No Test

Test reliability refers to the consistency of the test in measuring whatever it measures (Wiersma and Jurs, 2005). There are several ways of estimating reliability of a test, namely, parallel forms, test-retest, split-half, Kuder-Richardson and Cronbach alpha. The Cronbach's alpha (α) is the most commonly used (ibid.). Similarly, some studies on yes/no tests such as Shillaw (1996), Eyckmans (www, 2004) and Mochida and Harrington (2006) investigated internal consistency reliability of yes/no tests and the Cronbach's alpha (α) was used. The Cronbach's alpha is a measure of the degree to which the items are homogenous or consistent with each other (Richards and Schmidt 2002).

According to Chapter 3, there were two forms of the yes/no tests: form A and form B. Hence, the analysis of reliability was done in three ways: 1) the reliability of yes/no test Form A containing the first, the second, the third and the fourth 1000 most frequent word families respectively; 2) the reliability of yes/no test Form B containing the fourth, the third, the second and the first 1000 most frequent word families respectively and 3) both forms of the yes/no test. The results are shown in Table 4.2 below.

Table 4.2: Alpha Coefficients (α) for the Yes/no Test

	Numbers of Participants	**α**
Form A	40	0.968
Form B	35	0.955
Form A and Form B	75	0.962

The alpha coefficients derived from the three analyses in Table 4.2 were very close to 1. The alpha coefficients derived from the analysis ranged from 0.955 to 0.968. These were considered reliable because they were higher than the acceptable criterion of 0.70 as suggested in Fraenkel and Wallen (1993). Moreover, they were higher than the 0.82-0.87 reported in Beeckmans et al. (2001) although Beeckmans et al. had a significantly larger sample size (n = 488). This means that three forms of the yes/no tests all have high reliability. The high reliability reveals that the yes/no test tends to give the same results when it is given on different occasions or when it is used by different people (Richards and Schmidt, 2002). Therefore, the yes/no test is reliable enough to be used as a test for measuring participants' vocabulary size in the development of the vocabulary-based readability index and the validation of the index. However, it is too early to conclude that the test is an appropriate instrument for the study. We also need to investigate the concurrent and construct validity of the test.

4.1.3 Validation of the Yes/no Test

The yes/no test is sometimes criticized because the test aims to test the ability to recognize target words and understand their meaning but a test-taker is asked to only circle the words that they think they know. This leads to the problem with face validity, which refers to the degree to which the test appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of observers (Richards and Schmidt, 2002) because it does not look like a test of receptive vocabulary size. While other vocabulary size tests like Vocabulary Levels Test or VLT requires the test-taker to match definitions with target words, the yes/no test does not require him/her to do anything in order to show that they actually know the meaning of the target words that he/she circles. Therefore, it is questionable whether the test-taker actually knows the meaning of the words or he/she just recognizes that they have seen the words before. This section aims to examine the validity of the yes/no test in terms of concurrent and construct validity.

4.1.3.1 Concurrent Validity of the Yes/no Test

Concurrent validity is “the extent to which a test correlates with some other test that is aimed at measuring the same skill or with some other comparable measure of the skill being tested” (Richards and Schmidt, 2002). There have been several attempts to investigate concurrent validity of the yes/no test by comparing scores on the yes/no test with another test but they have produced mixed results

(Mochida and Harrington (2006). Meara and Buxton (1987) and Meara and Jones (1988) compared a yes/no test with a language proficiency test and found a moderately strong correlation where *r* was around 0.7. Similarly, Shillaw (1996) validated the yes/no test by comparing the scores on a yes/no test with scores on a multiple-choice proficiency test. However, he found a moderate correlation which was about 0.42 to 0.48. Cameron (2002) and Mochida and Harrington (2006) compared students' performance on the yes/no test with the VLT which is another kind of vocabulary size test. Cameron (2002) compared rankings of scores on yes/no test and VLT by using Spearman's rho in order to investigate the correlation between yes/no test and VLT level scores while Mochida and Harrington (2006) compared performance on identical items across all frequency levels in the yes/no test and VLT by Pearson *r*. The results of Cameron's study showed that there was no correlation between the two tests while Mochida and Harrington found out that there was a strong correlation between the two tests where *r* was around 0.8 and it is concluded that the scores on the yes/no test were a strong predictor of the performance on the VLT.

From the studies on the validation of yes/no test, we can see that the results were varied. Most of the studies on the concurrent validity of the yes/no test such as Meara and Buxton (1987); Meara and Jones (1988), Shillaw (1996) and Mochida and Harrington (2006) found positive correlations between scores on the yes/no test and scores on another vocabulary or vocabulary size test. However, the correlation coefficients are different. Meara and Buxton (1987), Meara and Jones (1988) as well as Mochida and Harrington (2006) report high correlation coefficients while Shillaw (1996) presents moderate ones. This is because some of these studies such as Meara and Buxton (1987); Shillaw (1996) and most of the items used in

Cameron (2002) have used independent measures of vocabulary knowledge so the content and format were confounded (Mochida and Harrington, 2006). Furthermore, only Cameron (2002) reports no correlation between the yes/no test and VLT. This might be because Cameron uses a different way of calculating correlations of scores. While other studies employ Pearson *r* in the calculation of correlations, Cameron changes the scores on the two tests into rankings and uses Spearman's rho to compare the rankings.

According to Phase 1 in Chapter 3, the concurrent validity was investigated in the pilot work of the yes/no test by using Pearson *r* in the same way as Anderson and Freebody (1983); Meara and Buxton (1987); Shillaw (1996); Eyckmans (www, 2004); and Mochida and Harrington (2006). Pearson *r* was used to investigate the correlations between the scores on the yes/no test and the scores on the translation test. Table 4.3 below presents the Pearson product-moment correlation coefficients between the yes/no test scores at each of the four levels (Yes/no Level 1, Yes/no Level 2, Yes/no Level 3 and Yes/no Level 4) and the scores on each of the four levels in the translation test (Translation Level 1, Translation Level 2, Translation Level 3 and Translation Level 4).

Table 4.3: Correlation Coefficients between the Yes/no Test and the Translation Test Scores

	Translation Level 1	**Translation Level 2**	**Translation Level 3**	**Translation Level 4**
Yes/no Level 1	.867**			
Yes/no Level 2		.868**		
Yes/no Level 3			.899**	
Yes/no Level 4				.686**

** Correlation is significant at level 0.01 (2 tailed)

The resulting correlation coefficients in Table 4.3 provide a measure of the concurrent validity. Most of the scores on the yes/no test in the table had strong correlations with scores on the translation test at $p < .01$ (two tailed) ranging from $r = .686$ for Level 4 to $r = .899$ for Level 3. This means that when the participants' scores on the yes/no test were high, their scores on the translation test were also high and when their yes/no test scores were low, their translation scores were also low.

From Table 4.3, three out of four r values were higher than the .7 which is the correlation coefficient between the yes/no test results and multiple-choice scores in Meara and Buxton (1987). These r values also mirrored the .84 reported in Anderson and Freebody (1983) for L1 speakers and the .78 to .87 which are the correlation coefficients between translation performances and yes/no test scores for $h - f$ reported in Mochida and Harrington (2006). It is more striking when the correlation coefficients in Table 4.3 are compared with the Cameron (2002) results. In that study, the highest correlation between yes/no and VLT level scores was .45 (Spearman's rho) and none were statistically significant.

So far, we have seen that the yes/no test in this pilot work has very high reliability and the extent to which the yes/no test, claming to test the ability to recognize target words and understand their meanings, correlates with the translation test, aiming to test the understanding of meaning, is high. In other words, the yes/no test has the consistency in measuring what it measures and it tends to correlate with another test that is aimed at measuring the same thing. Next, it is crucial to investigate construct validity in order to check whether the yes/no test measures the construct the test is supposed to measure.

4.1.3.2 Construct Validity of the Yes/no Test

As mentioned in 4.1.3.1, numerous studies on the yes/no test emphasize the concurrent validity. Unfortunately, the construct validity of the test has not been studied extensively (Eyckmans, www, 2004). Construct validity refers to the extent to which the test measures what it claims to measure (Brown, 1996). According to Beeckmans et al. (2001), the construct of the yes/no test is the ability to recognize the written form of target words and understanding their meanings. In order to investigate whether the yes/no test actually measures the ability to recognize the target words and understand their meanings or not, the answers on the yes/no test were compared with the responses on the VKS.

The comparison of the answers from both tests is presented in Table 4.4 below. According to Chapter 3, when the selected categories in the VKS were compared with the participants' performance on the yes/no test, there were four possible conditions: 1) the participants circled a target word in order to report that they know the meaning of the word and when they were asked to complete the VKS, they could translate the word correctly (Condition A); 2) the participants reported that they knew the meaning of the target word but they could not give accurate meaning or synonym of the target word (Condition B); 3) the participants reported in the yes/no test that they did not know the meaning of the word and they could not give accurate translation in the VKS (Condition C); and 4) the participants did not circle the target word but they could translate the word correctly (Condition D). Table 4.4 presents numbers of responses and percentage of responses where performance on the yes/no test and the VKS matched the aforementioned conditions.

Table 4.4: Numbers of Participants' Responses on the Yes/no Test and the VKS

		Condition A	Condition B	Condition C	Condition D
Level 1	**Number of participants**	267	55	98	44
	%	57.54	11.85	21.12	9.48
Level 2	**Number of participants**	115	88	234	27
	%	24.78	18.97	50.43	5.82
Level 3	**Number of participants**	97	78	266	23
	%	20.91	16.81	57.33	4.96
Level 4	**Number of participants**	61	60	327	16
	%	13.15	12.93	70.47	3.45
Total	**Number of participants**	540	281	925	110
	%	29.09	15.14	49.84	5.93

Table 4.4 shows that most of the participants' responses (925 times or 49.8429.09%) fell on Condition C. This means that most of them did not circle the target words in the yes/no test when they did not know the meaning of the words so they could not give correct translation or synonym to the target words in the VKS. Additionally, 540 responses or 29.09% of responses were yes responses to the target words in the yes/no test and the participants could give correct translation or synonym to the target words in the VKS. Surprisingly, there were 110 times that the participants reported no to a target word in the yes/no test but they could translate the word correctly in the VKS. This implies that some participants underestimated their own vocabulary knowledge while completing the yes/no test. In addition, there were 281 responses or 15.14% of responses that met Condition B. This shows that there were 281 times that the participants circled a target word in the yes/no test because they recognized the word in the test but they did not know its meaning so they could not translate the word correctly in the VKS. It implies that some of the participants

overestimated their own vocabulary knowledge while completing the yes/no test although pseudowords were added to the yes/no test. To study further why they overestimated their vocabulary knowledge, the participants' translation in the VKS was investigated.

The comparison of the responses on the yes/no test and the VKS also tells us why the participants said yes to some target words in the yes/no test although they did not know the meaning of the words. The main cause for this is that the participants have some problems with decoding. They were not so sure about the spelling of the words so they assumed that the target words were the words that they have known because they found the pronunciation or spelling familiar. For example, Participant A14 translated the word "heave" as "have", Participants A22 and A27 gave the definition of the word "greedy" as "glad", Participants A15, A16, A18, A20, A24, A25, A26 translated the word "tough" as "tongue", and Participant A30 gave the translation of the word "nasty" as "naughty".

Using the VKS to investigate the constructs of the yes/no test shows that it is likely that the yes/no test tests the ability to recognize the target words and to understand their meaning. However, we should be aware that it may not be the most accurate measure of vocabulary knowledge in terms of word meaning for Thai learners because around 20% of the responses did not represent the participants' actual vocabulary knowledge. This is higher than 11% that is the average tendency that test-takers would overestimate vocabulary knowledge in the yes/no test (Chall and Dale, 1950 quoted in Anderson and Freebody, 1981).

Besides the overestimation of vocabulary knowledge, there are some other factors affecting the participants' performance on the test. These include the language background e.g. Dutch learners of French in Eyckmans (www, 2004) had higher false alarm rates than the participants whose L1 is not based on Latin alphabet in Mochida and Harrington (2006), the proficiency of the learner (Meara, 1996) i.e. test validity generally increases as proficiency level increases (Mochida and Harrington, 2006), length of the test, the participants' motivation, honesty, confidence and decoding skills. While acknowledging the role of these factors can play, the yes/no test has been recognized as a valid and potentially useful estimate of vocabulary (Read, 2000).

According to McNamara (2000), there is no best test method because each method has both advantages and disadvantages. The yes/no test is a test method and it does not seem to be the best and the most accurate test of the ability to understand word meaning but because of its merits, it has been used to test vocabulary size of native speakers of English, second language learners and foreign language learners.

4.2 The Results of the Pilot Work of the Reading Comprehension Test

The second section of this chapter involves the presentation of the results of the pilot work of the reading comprehension test. It consisted of three main parts: 1) face validity of the reading comprehension test; 2) test reliability and 3) item difficulty and item discrimination.

4.2.1 Face Validity of the Reading Comprehension Test

According to Hughes (1989; 2003); Alderson, Clapham and Wall (1995); Johnson (2001); and Brown (2005), there are three common types of validity which are content validity, criterion-related validity including concurrent and predictive validity and construct validity. Among these types, content validity seems to be the first one to establish the validity of the test or whether the test measures what it aims to measure through the assessment of experts (Wiersma and Jurs, 2005). In order to do this, we need to insure that the experts really are experts (Brown, 2005).

Since the study aims to devise the vocabulary-based readability index for teachers and students to select texts, teachers and students are the only two groups involving in the study. It will be great to ask teachers who have experience in writing reading tests to decide whether the test measures what it is supposed to measure. However, it is questionable whether the teachers really are experts. In order to avoid any misconception of the term "experts", the term "face validity" was used in the investigation although the procedures of investigating content validity were followed. Face validity refers to the degree to which the test appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of observers (Richards and Schmidt, 2002). A test is said to have face validity if it looks as if it measures what it is supposed to measure (Hughes, 1989; 2003).

Like content validity, face validity in the present study was investigated by the questionnaire with 3-point scale (see Appendix E). The teachers or participants in Group 4 were asked to look through each test item or question in the reading comprehension test and rate whether they agree that each item appears to measure the purpose of the test which is to test the participants' ability to scan and paraphrase specific information. Then they were supposed to rate on the 3-point scale which valued 1 (agree), 0 (unsure) and -1 (disagree). The results obtained from the questionnaire (see Appendix M) were analyzed by mean (see 3.2.3.1). The results are presented in Table 4.5 below.

Table 4.5: The Mean of Participants' Responses in the Questionnaire for Investigating Face Validity of Reading Comprehension Test

Items	Mean	Items	Mean
1	1.0	13	0.6
2	1.0	14	0.8
3	1.0	15	1.0
4	1.0	16	1.0
5	1.0	17	1.0
6	0.8	18	1.0
7	0.6	19	0.6
8	1.0	20	1.0
9	0.8	21	1.0
10	0.8	22	1.0
11	1.0	23	1.0
12	1.0	24	1.0

According to 3.2.3.1, if the mean was 0.5 to 1.0, it meant that the teachers agreed that the test looked good as a test of reading comprehension in terms of scanning and paraphrasing. On the other hand, if the mean was below 0.5, it meant that the test did not look good and did not have face validity. From Table 5.5, we can see that the minimum mean was 0.6 (items 7, 13 and 19) and the maximum value was 1. In addition, the mean of lecturers' responses on the 24 items were higher than 0.5, this meant that the test looked good as a test of reading comprehension in terms of scanning and paraphrasing.

4.2.2 Reliability of the Reading Comprehension Test

Test reliability refers to the consistency of the test in measuring whatever it measures (Wiersma and Jurs, 2005). According to Phase 2 of Chapter 3, test reliability of the reading comprehension test was investigated by Cronbach's alpha (α). The alpha coefficient derived from the analysis was 0.78 which was considered reliable because it was higher than the acceptable criterion of 0.70 as suggested in Fraenkel and Wallen (1993). Therefore, it is likely to say that this reading comprehension test has reliability and it tends to consistently test reading comprehension in terms of scanning and paraphrasing.

4.2.3 Item Analysis

The results of item analysis of the reading comprehension test are presented in the table below.

Table 4.6: Results of Item Analysis of the Reading Comprehension Test

Item Number		Item Difficulty	Item Discrimination	Remark
Passage One	1	0.76	0.30	Appropriate
	2	0.71	0.57	Appropriate
	3	0.62	0.24	Appropriate
	4	0.71	0.83	Appropriate
	5	0.57	0.27	Appropriate
	6	0.57	0.59	Appropriate
Passage Two	7	0.56	0.54	Appropriate
	8	0.49	0.24	Appropriate
	9	0.53	0.46	Appropriate
	10	0.52	0.50	Appropriate
	11	0.50	0.35	Appropriate
	12	0.48	0.30	Appropriate
Passage Three	13	0.47	0.49	Appropriate
	14	0.44	0.35	Appropriate
	15	0.42	0.49	Appropriate
	16	0.43	0.35	Appropriate
	17	0.44	0.51	Appropriate
	18	0.42	0.54	Appropriate

Table 4.6: Results of Item Analysis of the Reading Comprehension Test (Cont.)

Item Number		Item Difficulty	Item Discrimination	Remark
Passage Four	19	0.35	0.30	Appropriate
	20	0.38	0.43	Appropriate
	21	0.42	0.54	Appropriate
	22	0.40	0.38	Appropriate
	23	0.40	0.46	Appropriate
	24	0.38	0.43	Appropriate

Table 4.6 shows that all of the items were acceptable because the values of item difficulty and item discrimination reach the aforementioned criteria. However, the values of the difficulty level of items 19, 20 and 24 and power of discrimination of items 3, 5 and 8 are not very high. These low values might be caused by the following reasons.

For the difficulty level, the participants' performance on each of the four sections in the test may be dependent to their vocabulary knowledge. If most of the participants who have a large vocabulary size on the fourth 1000 most frequent word families work on the test, they tend to get high scores on the last section containing a reading passage with 95% of tokens from the first, the second, the third and the fourth 1000 most frequent word families (Laufer, 1989). When most of them can answer the questions in the section correctly, the level of difficulty will show that the test items or questions are too easy for the participants. There are two possible explanations for this: 1) the test items were not well written so the participants needed to make random guesses; and 2) most of the participants get the correct answers because they know 95% of tokens so they can comprehend the passage (Laufer, 1989).

The low level of discrimination may be affected by two possible reasons: 1) the participants who have large vocabulary size seem to be able to answer most of the questions in the test correctly. This is supported by Laufer (1992) reporting that readers who know the first 3000 most frequent word families can read unsimplified texts and Nation (1990, 2001) and Hirsh and Nation (1992) stating that readers who know the first 5000 most frequent word families can comprehend unsimplified texts. If most of the participants have similar vocabulary size, they will be able to answer most of the questions correctly. When all of them answer the questions correctly, the level of discrimination will be low. This low level of discrimination may not mean that the test item or question is not good enough to discriminate good and poor students but it may imply that most of the participants have similar vocabulary size; and 2) the test items with low level of discrimination may be ambiguously written (Office of Educational Assessment, University of Washington, www, 2005).

Although reasons of low level of difficulty and power of discrimination are given, the researcher was aware that the test items with low values for difficulty level and power of discrimination might be ambiguously worded so the language use in the aforementioned test items was examined in order to make sure that all of them are well worded.

4.3 Summary

This chapter described how the results from the pilot work of yes/no test and the pilot work of reading comprehension test. The yes/no test is claimed to test the ability to recognize target words and understand their meanings (Beeckmans et al., 2001). However, it tends to test only word recognition because it presents the test-takers or participants a set of target words and pseudowords and asks them to report whether they know the words or not. Fortunately, the correlation coefficients between the scores on the yes/no test and the scores on the translation test were quite high. This shows that yes/no test correlates with the translation test which aims to measures the understanding of meanings. Moreover, Cronbach's alpha coefficients were very high ranging from 0.955 to 0.968. This shows that the yes/no test consistently tests what it is measuring. Additionally, the comparison of the answer on the yes/no test and the selected categories in the VKS revealed that most of the participants did not circle the target words when they did not know the words and 29.09% of them actually knew the meanings of the words that they circled**.** There were very few of them who recognized the words but could not translate the words correctly and who did not recognize the words but could give correct translation.

For the reading comprehension test, the mean or lecturers' average responses in the questionnaire for investigating face validity revealed that the test appeared to measure the ability to scan and paraphrase specific information. For test reliability, the Cronbach's alpha coefficient (0.78) reveals that the reading comprehension test seemed to have consistency in measuring the ability to scan and paraphrase specific information. As a result, the reading comprehension test could be used in the development of the vocabulary-based readability index (Phase 3) and the validation of the index (Phase 4) following data collection procedures and data analysis presented in Chapter 3.

CHAPTER 5

VOCABULARY-BASED READABILITY INDEX

The main purpose of this chapter is to present the findings of the study in response to the research question "To what extent would a purely vocabulary-based readability index be valid?". This chapter is organized into three main sections. The first section involves the investigation of the effects of independent variables (LFP, vocabulary size and TSVK) used for devising the vocabulary-based readability index on the dependent variable (reading comprehension scores) by three-way ANOVA. The second section deals with the development of the vocabulary-based readability index by simple regression analysis. The last section presents the validation of the index in terms of face validity, concurrent validity and predictive validity by Spearman's rank correlation or Spearman's rho and Pearson product-moment correlation or Pearson *r*.

5.1 An Investigation of the Effects of Lexical Frequency Profile (LFP), Vocabulary Size and Text-specific Vocabulary Knowledge (TSVK) on Reading Comprehension Scores

This part aims to investigate the effects of the three independent variables which are LFP, vocabulary size and TSVK on the dependent variable which is reading comprehension scores. It starts with descriptive statistics including mean and standard deviation (*SD*) so as to give an overview of the reading comprehension scores. Then a three-way ANOVA is used to investigate whether there is a significant difference of mean reading scores when each of the independent variables is varied and whether the interactions between the independent variables have any effect on reading scores. When there was a significant difference found in the result of three-way ANOVA, Scheffé test, a post hoc test, was used to indicate which of the differences between particular pairs of mean scores are contributing to the overall differences.

Descriptive statistics was employed in order to get an overview of participants' reading scores on each of the four reading passages in the reading comprehension test. The mean reading scores of 102 participants who were asked to complete all of the four reading passages in the reading comprehension test (see Appendix D) and the standard deviation are shown in the table below. The raw scores are presented in Appendix N.

Table 5.1: Descriptive Statistics for the Participants' Reading Scores on the Four Reading passages in the Reading Comprehension Test

Reading Passage	N	Mean	*SD*
One	102	4.34	1.301
Two	102	3.44	1.638
Three	102	2.65	1.657
Four	102	2.25	1.325

Table 5.1 shows that the mean reading score on Passage One (mean = 4.34, *SD* = 1.301) was the highest while the mean score on Passage Four was the lowest (mean = 2.25, *SD* = 1.325). This is because according to Chapter 3, 95% of tokens in Passage One are from the first 1000 most frequent word families so this passage tends to be the easiest one because most of the tokens in the passage are high frequent

words and readers or students tend to be familiar with them. On the other hand, 95% of tokens in Passage Four are from the first, the second, the third and the fourth 1000 most frequent word families. There are fewer high frequency words in Passage Four than the other three so Passage Four tends to more difficult than the first three reading passages.

The reading comprehension scores which are the dependent variable were arranged into groups based on the three independent variables: 1) LFP; 2) TSVK; and 3) vocabulary size. These arrangements were necessary in order to find out whether the dependent variable or the reading scores on the four texts varied significantly according to the independent variables which are vocabulary sizes, TSVK, and LFP by using three-way ANOVA.

The reading scores were categorised into four groups based on LFP: 1) the reading scores on reading comprehension questions from Passage One with 95% of tokens from the first 1000 most frequent word families; 2) the scores on questions on Passage Two consisting of 95% of tokens from the first and the second 1000 most frequent word families; 3) the scores on questions from Passage Three with 95% of tokens from the first, the second and the third 1000 most frequent word families; 4) the reading scores on questions from Passage Four consisting of 95% of tokens from the first, the second, the third and the fourth 1000 most frequent word families. Since each of the four groups of reading scores arranged by LFP were scores on questions from each reading passage, the mean reading scores of all of the four groups arranged by LFP and the standard deviation were identically the same as Table 5.1.

For the scores on the yes/no test, the score 0.7 or 70% is generally used for criterion-reference marking as a benchmark showing that students who get 70% or more have performed well throughout an academic term or have achieved the learning objectives of the course in most of the high schools such as Satriwittaya School and Suankularbwittayalai School and most universities in Thailand, namely Kasetsart University and Silpakorn University. In other words, seventy percent of scores on assignments and tests assigned in a course is considered good in most of the high schools in Thailand. Seventy percent is normally represented by number 3 in high schools and grade B in universities. Hence, this score was used as a benchmark for judging whether the participants know enough vocabulary on a particular frequency band to say that they know the words in the frequency range. Based on the scores on the yes/no test, the reading scores were arranged into five groups: 1) the reading scores of participants whose vocabulary size was below 1,000 or of those whose score on the first 1000 most frequent word families in the yes/no test was lower than 0.7; 2) reading scores of the participants whose vocabulary level was 1000 or of those whose score on the first 1000 most frequent word families in the yes/no test was 0.7 or higher; 3) reading scores of those whose level was 2000 or of those whose score on the first and the second 1000 most frequent word families in the yes/no test was 0.7 or more; 4) scores of those whose vocabulary level was 3000 or of those whose score on the first, the second and the third 1000 most frequent word families was 0.7 or higher; and 5) scores of the participants whose vocabulary level was 4000 or of those whose score on the first, the second, the third and the fourth 1000 most frequent word families in the yes/no test was 0.7 or higher.

In order to provide an overview of the reading scores in each of the five groups, descriptive statistics of the five groups of reading scores arranged by vocabulary size were used. The number of selected real words, pseudowords and the scores are presented in Appendix O. The mean reading scores from all of the five groups and standard deviation are presented below.

Table 5.2: Descriptive Statistics for the Participants' Reading Scores Arranged by Vocabulary Size

Groups of Reading Scores	N	Mean	*SD*
Group 1	144	1.97	1.226
Group 2	112	3.04	1.536
Group 3	56	3.84	1.276
Group 4	64	4.48	1.039
Group 5	32	5.28	1.687

Table 5.2 shows that the mean reading score of the participants whose vocabulary size was below 1000 (Group 1) was the lowest (mean = 1.97, *SD* = 1.226) while the mean score of the participants whose vocabulary size was 4000 (Group 5) was the highest (mean = 5.28, *SD* = 1.687). It reveals that the mean reading score increases when the participants' vocabulary size is larger.

The reading scores were also arranged into three groups on the basis of TSVK following the findings from Laufer (1989): 1) the reading scores of participants whose TSVK is 95% or higher; 2) the scores of those whose TSVK is between 94% and 90% and 3) the reading scores of those whose TSVK is 89% or lower.

Descriptive statistics including mean and *SD* was carried out in order to get an overview of the reading scores in each of the three groups. The results are presented in the table below.

Table 5.3: Descriptive Statistics for the Participants' Reading Scores Arranged by TSVK

Groups of Reading Scores	N	Mean	*SD*
Group 1	300	3.82	1.391
Group 2	108	1.37	0.982
Group 3	0	-	-

Table 5.3 shows that TSVK of most participants (N = 300) was 95% or higher and the mean reading score of these participants (Group 1) was the highest (mean = 3.82, *SD* = 1.391). The mean reading score of the participants whose TSVK was between 94%-90% (Group 2) was lower (mean = 1.37, *SD* = 0.982). Additionally, none of them had TSKV that was 89% or lower.

To compare the mean reading scores arranged by 1) LFP, 2) vocabulary size and 3) TSVK, the three-way ANOVA was employed. This statistical method was selected for three reasons: 1) there were three independent variables which are LFP, vocabulary size and TSVK; 2) three means were compared; and 3) the data under comparison were interval and normally distributed. Seven null hypotheses (H_0) and alternative hypotheses (H_1), which were tested by the three-way ANOVA, were set as follows:

1) H_0: There is no difference of the mean reading scores when the LFP is varied.
   H_1: There is some difference of the mean reading scores when the LFP is varied.
2) H_0: There is no difference of the mean reading scores when the vocabulary size is different.
   H_1: There is some difference of the mean reading scores when the vocabulary size is different.

3) H_0: There is no difference of the mean reading scores when the TSVK is varied.

   H_1: There is some difference of the mean reading scores when the TSVK is varied.

4) H_0: There is no interaction between LFP and vocabulary size.

   H_1: There is an interaction between LFP and vocabulary size.

5) H_0: There is no interaction between LFP and TSVK.

   H_1: There is an interaction between LFP and TSVK.

6) H_0: There is no interaction between vocabulary size and TSVK.

   H_1: There is an interaction between vocabulary size and TSVK.

7) H_0: There is no interaction between LFP, vocabulary sizes and TSVK.

   H_1: There is an interaction between LFP, vocabulary size and TSVK.

Before we can use the three-way ANOVA to test the hypotheses, it is important to investigate whether the reading comprehension scores are normally distributed or not. Below are a bar chart and a normal curve showing how the reading comprehension scores are distributed.

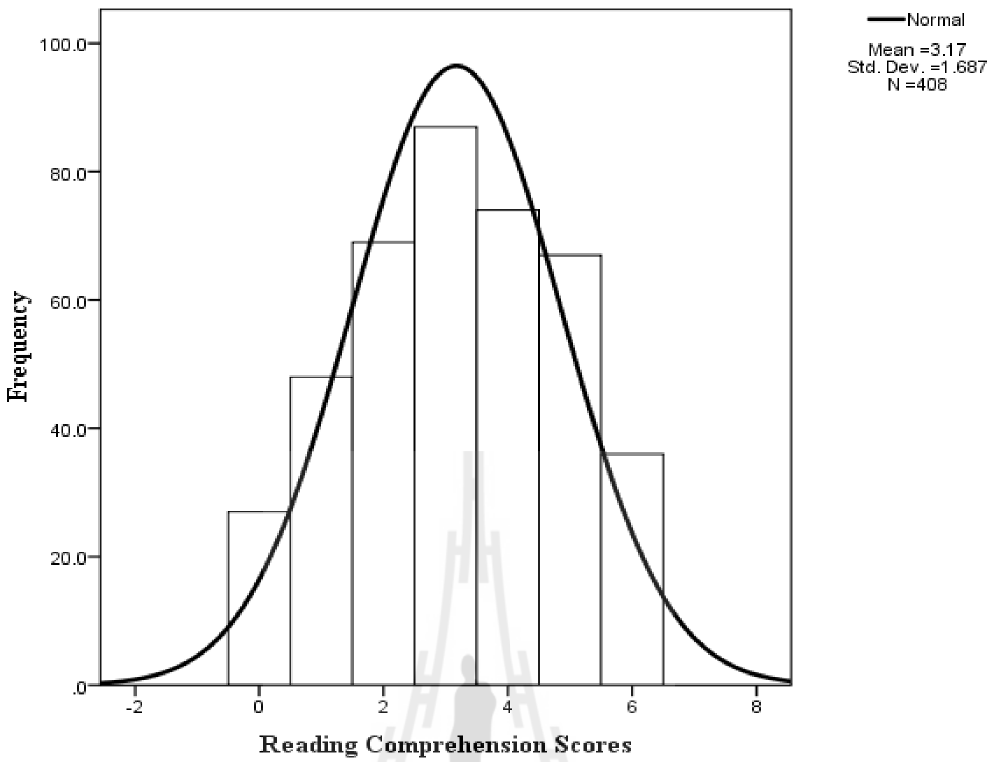


Figure 5.1: Normal Curve of Distribution of Reading Comprehension Scores

The normal curve in Figure 5.1 is shaped like a bell that peaks in the middle and is perfectly symmetrical. This shows that the distribution of data seems to be normal.

The three-way ANOVA were calculated in order to test the hypotheses above. The level of significance for testing the hypotheses was set at .05. The results of three-way ANOVA are shown below.

Table 5.4: The results of three-way ANOVA

Source of Variation	*SS*	df	*MS*	*F*	*p-value*
Main effects					
LFP	53.888	3	17.963	21.880	.000
vocabulary size	119.060	4	29.765	36.256	.000
TSVK	32.458	1	32.458	39.537	.000
Two-way interactions					
LFP*vocabulary size	19.834	12	1.653	2.013	.022
LFP*TSVK	4.878	3	1.626	1.980	.116
Vocabulary size*TSVK	1.034	3	0.345	0.420	.739
Three-way Interactions					
LFP*vocabulary size*TSVK	4.000	3	1.333	1.624	.183
Error	310.325	378	0.821		
Total	5262.000	408			

Table 5.4 shows the results of the three-way ANOVA. It was found that for the main effects, the first three null hypotheses, which are 1) there is no difference of the mean reading scores when the LFP is varied (F(3, 378) = 21.880, p < .01), 2) there is no difference of the mean reading scores when the vocabulary size is varied (F(4, 378) = 36.256, p < .01) and 3) there is no difference of the mean reading scores when the TSVK is varied (F(1, 378) = 39.537, p < .01), were rejected and we needed to accept the alternative hypotheses. This means that the mean scores of more than two groups of the participants' reading scores were significantly different when each of the three independent variables which are LFP, vocabulary size and TSVK were varied. Additionally, for two-way interactions, only one out of three null hypothesis, which is 4) there is no interaction between LFP and vocabulary sizes (F(12, 378) = 2.013, p < .05) was rejected and the alternative hypothesis was then accepted. This reveals that there were some interactions between LFP and vocabulary size (LFP*vocabulary size) at the significant level at .05. This reveals the interactions between LFP and vocabulary size had some effect on reading comprehension scores.

For the other three null hypotheses (hypotheses 5, 6 and 7), it was found that the null hypothesis 5, which is there is no interaction between LFP and TSVK, was accepted ($F(3, 378) = 1.980$, $p = .116$) because the p-value (.116) was greater than .05 which is the level of significant set before. This means that the reading scores were not affected by the interactions between LFP and TSVK. Similarly, the null hypothesis 6) which is there is no interaction between vocabulary size and TSVK ($F(3, 378) = 0.420$, $p = .739$). It shows that the interactions between LFP and TSVK as well as the interactions between vocabulary size and TSVK did not have any effect on reading comprehension scores. Moreover, the null hypothesis 7, which is there is no interaction between LFP, vocabulary size and TSVK, was also accepted ($F(3, 378) = 1.624$, $p = .183$). This reveals that the interactions between the three variables did not have any effect on the reading scores.

When there were significant differences as the results of three-way ANOVA among groups of reading scores arranged by LFP, vocabulary size and TSVK, it is crucial to execute Scheffé tests, a post hoc test, to pinpoint which of the differences between particular pairs of mean scores are contributing to these overall differences. The differences between pairs of mean reading scores arranged by LFP and vocabulary size were calculated and the results are presented in Table 5.5 and Table 5.6 respectively. The post hoc test was not performed for TSVK because there were only two groups.

Table 5.5: Results from the Scheffé Test on Differences of Pairs of Mean Reading Scores Arranged by LFP

Group	N	Subset			
		1	**2**	**3**	**4**
4	102	2.25			
3	102		2.65		
2	102			3.44	
1	102				4.34
Sig.		1.000	1.000	1.000	1.000

The post-hoc Scheffé test was calculated to identify where the significant difference occurred among the groups. From Table 5.5, the presentation of mean scores in the subset and the significance value (1.000) reveals that the average score on all of the four groups were significantly different from each other. The average score on the reading passage with 95% of tokens from the first 1000 most frequent word families (4.34) was significantly higher than the average scores on other reading passages and the average score on the reading passage with 95% of tokens from the first, the second, the third and the fourth 1000 most frequent word families (2.25) was significantly lower than other average scores. It can be noted that the participants' reading scores tend to significantly decrease when the number of high frequency words decreases. In other words, the participants tend to find the texts with a lot of high frequency words easier than the texts with fewer high frequency words. This implies that LFP has some effect on how difficult a reading passage is for participants so LFP will be included in the development of the vocabulary-based readability index.

Table 5.6: Results from the Scheffé Test on Differences of Pairs of Mean Reading Scores Arranged by Vocabulary Size

Group	**N**	**Subset**				
		1	**2**	**3**	**4**	**5**
1	144	1.97				
2	112		3.04			
3	56			3.84		
4	64				4.48	
5	32					5.28
Sig.		1.000	1.000	1.000	1.00	1.00

Table 5.6 shows that the mean scores of all of the five groups were significantly different. Moreover, the Scheffé test reveals that the mean reading score of the participants whose vocabulary size was 4000 (5.28) was significantly higher than the mean scores of other groups and the mean reading score of the participants whose vocabulary size was below 1000 (1.97) was significantly lower than others. This explains that the participants whose vocabulary size is small tend to have lower reading scores than the ones with a large vocabulary size and it is likely that the participants' reading scores increase when they have a larger vocabulary size.

Results of the descriptive analyses and the three-way ANOVA reveal that reading scores tend to be different when the text-based variable (LFP) and the reader-based variables (vocabulary size and TSVK) are varied. These results support the fact that 1) the percentage of basic vocabulary a text contains can predict text difficulty (Laufer, 1997); 2) knowing more words would assure better comprehension of reading passages (Horst, Cobb and Meara, 1998); and 3) knowledge of the vocabulary in a text (TSVK) is one of the many factors affecting reading (Hirsh and Nation, 1992). Therefore, it looks promising to investigate whether these variables can be used to indicate text difficulty in the vocabulary-based readability index. The development and the validation of the index will be presented in 5.2 and 5.3 respectively.

5.2 The Development of Vocabulary-based Readability Index

As mentioned in Chapter 2, there are two aspects of text difficulty: 1) the difficulty of text or texts and 2) how each participant finds a text difficult. In order to do these, it is important to use the only text-based variable which is LFP to indicate difficulty of texts and the reader-based variables which are vocabulary size and TSVK to predict reading comprehension scores or reader's ability to understand a text. According to the framework for developing the vocabulary-based readability index in Chapter 2, there are three components of the index: 1) indication of text difficulty by LFP; 2) prediction of reading comprehension at different vocabulary size; and 3) prediction of reading comprehension at different TSVK.

The first component of the vocabulary-based readability index concerns the use of LFP as an indicator of text difficulty because it is the only text-based variable in the study (see 5.2.1). The second and the third components of the index involve the prediction of levels of difficulty of texts for individual students. The levels of difficulty will be predicted by vocabulary size (see 5.2.2) and TSVK (see 5.2.3).

5.2.1 Indication of Text Difficulty by Lexical Frequency Profile

When we look at the results of the three-way ANOVA for LFP shown in Table 5.4 again, we can see that the three-way ANOVA provided a comparison of the actual reading scores across the different LFP and the results showed that there were significant differences among the reading scores on the four reading passages with different LFP at .01 level ($F(3, 378) = 21.880$, $p < .01$). In addition, the results from the Scheffé test in Table 5.5 revealed that the participants' reading scores tend to decrease when the proportion of low frequency words increases. This ensures that the percentage of basic vocabulary a text contains can predict text difficulty (Laufer, 1997) and difficulty of advanced texts can also be predicted by the increased percentage of low frequency words (Hacquebord and Stellingwerf, 2007). These results show great promise for using LFP to indicate text difficulty.

However, there are three possible indications of text difficulty by LFP. It is important to discuss each of them in detail and choose the most suitable one for the study by using Pearson *r*. One is based on Laufer (1989) stating that readers who know 95% of tokens in a text can comprehend that text. Hence, we should use 95% as the benchmark judging the number of word lists or frequency bands (e.g. the first 1000 most frequent word families, the second 1000 most frequent word families, the third 1000 most frequent word families) that is necessary for text comprehension. By way of illustration, let us look at proportion of low and high frequency words from Text 7 and Text 9 and how the difficulty of these two texts is indicated.

Table 5.7: Lexical Frequency Profile of Text 7 and Text 9 in Appendix P

Word List/Frequency Band	Tokens/%	
	Text 7	Text 9
the first 1000	100/73.53	71/64.55
the second 1000	24/17.65	19/17.27
the third 1000	3/2.21	7/6.36
the fourth 1000	3/2.21	3/2.73
the fifth 1000	1/0.74	0/0.00
the sixth 1000	0/0.00	2/1.82
the seventh 1000	0/0.00	3/2.73
the eighth 1000	0/0.00	1/0.91
the ninth 1000	0/0.00	1/0.91
the tenth 1000	0/0.00	0/0.00
the eleventh 1000	2/1.47	1/0.91
the twelfth 1000	0/0.00	0/0.00
the thirteenth 1000	0/0.00	0/0.00
the fourteenth 1000	0/0.00	0/0.00
the fifteenth 1000	2/1.47	0/0.00
the sixteenth 1000	0/0.00	0/0.00
Off-list	1/0.74	2/1.82

From Table 5.7, we can see that 95% of tokens in Text 7 come from four frequency bands which are the first, the second, the third and the fourth 1000 most frequent word families while 95% of tokens in Text 9 come from seven frequency bands. If we use this as the benchmark for indicating text difficulty, we will take the numbers of frequency bands that make up 95% of tokens in texts into consideration. The text with 95% of tokens from fewer frequency bands seems to be easier than the text with 95% of tokens from more frequency bands. Hence, Text 7 seems to be easier than Text 9.

Another way of indicating text difficulty by LFP is to look at the number of tokens or percentage of tokens from the first, the second and the third 1000 most frequent word families because readers who know 3000 word families covering 95% of most texts can comprehend unsimplified texts (Nation and Waring, 1997). In order to apply this notion to the indication of text difficulty, we need to combine all of the tokens from the first, the second and the third 1000 most frequent word families and calculate the percentage. Then if we want to know how difficult the text is, we should look at the percentage of tokens from the three frequency bands. If the percentage is very close to 95, this means that the text seems to be readable for the ones whose vocabulary size is 3000. If we want to compare the difficulty of two or more texts, we should compare percentage of a text against others. In this case, we should compare the percentage of tokens in each text from the first three frequency bands. The text with the higher percentage of tokens tends to be easier. Let us look at Table 5.7 again. From this table, we can see that Text 7 has 136 tokens and 127 tokens are from the first three frequency bands which make up 93.39% while Text 9 has 97 out of 110 tokens which make up 88.13%. This shows that Text 7 seems to be easier than Text 9.

The other way to indicate text difficulty by LFP is to look at the percentage of tokens from the first frequency band (the first 1000 most frequent word families) because these words families most frequently occur in everyday life. Readers tend to be familiar with them so it is likely that a text with more words from this frequency band tends to be easier. For instance, the percentage of tokens in Text 7 from the first frequency band is 73.53 and the percentage of tokens in Text 9 from the first frequency band is 64.55. When Text 7 has more high frequency words than Text 9, it implies that Text 7 tends to be easier than Text 9.

So far, we have seen that there is more than one way of using LFP to indicate text difficulty. It is important to select the most suitable one for the study. In order to do so, we need to compare them to some readability formulas which are considered as an objective measure of readability. The three different uses of LFP to indicate text difficulty and two readability formulas which are Flesch readability formula and Flesch-Kincaid readability test were used to indicate text difficulty of 30 texts. Flesch reading ease score derived from Flesch readability formula and Flesch-Kincaid grade level obtained from Flesch-Kincaid readability test were selected because they are prominently used by most material designers and teachers (RFP Evaluation Centers, www, 2008). They are also used as the measures of readability of documents in Microsoft Word (Ownby, 2005) which are ubiquitously operated.

These texts were randomly selected from Google top searches on the Internet such as Pollick (www, 2010) and Ruiz (www, 2008). The texts were also randomly selected from reading practice books such as Brown and Hood (2004) and Educational Testing Service (2007c). The length of the texts ranged from 89 words to 507 words (see Appendix P). They were then analyzed by the three different uses of LFP, Flesch readability formula and Flesch-Kincaid readability test. For the percentage of tokens from the first three frequency bands, the percentage of tokens from the first frequency band and the results or reading ease scores of Flesch readability formula, the results range from 0 to 100. Higher percentage or scores indicate texts that are easier to read and lower numbers mark texts that are more difficult to read. Conversely, for the number of frequency bands making up 95% of tokens in a text, the results range from 1 to 16. High numbers of frequency bands indicate texts that are difficult to read and fewer numbers of frequency bands mark texts that are easier to read. Similarly, for the results or American grade level from Flesch-Kincaid readability test, the possible results analyzed by Micorsoft Word 2003 range from 0 to 12. Texts that are predicted to be suitable for higher level tend to be more difficult and texts that are said to be suitable for lower level seem to be easier.

The results or raw scores from these five methods are presented in Appendix Q. The correlations between results from all of the aforementioned methods of indicating text difficulty by LFP and the two readability formulas were investigated by Pearson product-moment correlation coefficient or Pearson *r* with the assistance of SPSS. The results are presented below.

Table 5.8: Pearson's Correlation Coefficients between Results of Text Difficulty Indicated by LFP and Two Readability Formulas (N = 30)

	Flesch Readability Formula	**Flesch-Kincaid Readability Test**
Number of Frequency Bands Making Up 95% of Tokens	r = -0.187 p = 0.322	r =0.328 p = 0.077
Percentage of Tokens from the First Three Frequency Bands	r =0.364* p = 0.048	r =-0.480** p = 0.007
Percentage of Tokens from the First Frequency Band	r =0.375* p = 0.041	r =-0.309 p = 0.097

* Correlation is significant at the 0.05 level (two-tailed)
** Correlation is significant at the 0.01 level (two-tailed)

From Table 5.8, the results of Pearson *r* show that the number of frequency bands making up 95% of tokens did not correlate with the results of Flesch readability formula (r = -0.187, p = 0.322) and Flesch-Kincaid readability test (r = 0.328, p = 0.077) because the *p*-values were greater than 0.05 which is the significance level accepted by social scientists. Moreover, there was no correlation between the percentage of tokens from the first frequency band and the results from Flesch-Kincaid readability test (r = -0.309 p = 0.097). It was also found that the percentage of tokens from the first three frequency bands (the first, the second and the third 1000 most frequent word families) was positively correlated with the results of Flesch readability formula with a correlation coefficient of r = 0.364 (p = 0.048) which is significant at $p < 0.05$ and the percentage of tokens from the first frequency band positively correlated with the results of Flesch readability formula with a correlation coefficient of $r = 0.375$ ($p = 0.041$) which is significant at $p < 0.05$. This positive correlation shows that as the percentage of tokens from the first three frequency bands increases, the results from Flesch readability formula also increase. Likewise, as the percentage of tokens from the first three frequency bands decreases, the results from Flesch readability formula also decrease. That is to say, both of them predicted difficulty of the thirty texts in the same way. When a text is easy, the percentage of tokens from the first three frequency bands and the results of Flesch readability formula were both high and when a text is difficulty, both of them were low. On the other hand, the percentage of tokens from these three frequency bands was negatively correlated with the results of Flesch-Kincaid readability test with a correlation coefficient of $r = -0.480$ which is highly significant at $p < 0.01$. This negative correlation reveals that when the percentage of tokens from the first three frequency bands increases or decreases, the results of Flesch-Kincaid readability test will change in the opposite direction. That is because they predict text difficulty in different directions (see above). That is to say, when a text is said to be easy by the percentage of tokens from the first three frequency bands, the percentage is high and when the text is said to be easy by Flesch-Kincaid readability test, the result or predicted American grade level is low. When a text is indicated to be difficult to read by the percentage of tokens from the first three frequency bands, the percentage is low and when a text is marked as difficult to read by Flesch-Kincaid readability test, the predicted American grade level is high.

When we look at the significant correlations, we can see that the correlation coefficients obtained from the correlation between the percentage of tokens from the first three frequency bands and the results of Flesch readability formula ($r = 0.364$, $p < 0.05$); the correlation between the percentage of tokens from the first three frequency bands and the results of Flesch-Kincaid readability test ($r = -0.480$, $p < 0.01$); and the correlation between the percentage of tokens from the first frequency band and the results from Flesch readability formula ($r = 0.375$, $p < 0.05$) ranged from -0.480 to 0.375. According to Roscoe (1975), the correlation coefficients of 0.30 to 0.70 show a moderate correlation, those below 0.30 indicate a low correlation and those larger than 0.70 mean a high correlation. This shows that although both percentage of tokens from the first three frequency bands and the percentage of tokens from the first frequency band significantly correlate to the results of Flesch readability test and the correlation between the percentage of tokens from the first three frequency bands and the results of Flesch-Kincaid readability test is highly significant, the correlations are moderate. It is because the predictions of text difficulty by LFP, Flesch readability formula and Flesch-Kincaid readability test are based on different measures. Flesch readability formula and Flesch-Kincaid readability test are based on the same core measures which are word length and sentence length while LFP is purely based on proportion of high frequency words.

Although Table 5.8 shows that the percentage of tokens from the first three frequency bands looks more promising than the other two because it significantly correlates with the results of Flesch readability formula and Flesch-Kincaid readability test, the correlation coefficients are not very high. In order to look for more evidence, the percentage of tokens from the first three frequency bands was compared with the percentage of tokens from the first frequency band and the number of frequency bands making up 95% of tokens because as mentioned earlier, all of them are based on the same measure. The raw data are presented in Appendix Q and the results of Pearson *r* are presented below.

Table 5.9: Pearson's Correlation Coefficients between Results of Text Difficulty Indicated by LFP (N = 30)

	Percentage of Tokens from the First Frequency Band	**Percentage of Tokens from the First Three Frequency Bands**
Number of Frequency Bands Making Up 95% of Tokens	$r = -0.453$* $p = 0.012$	$r = -0.683$** $p = 0.000$
Percentage of Tokens from the First Three Frequency Bands	$r = 0.709$** $p = 0.000$	

* Correlation is significant at the 0.05 level (two-tailed)
** Correlation is significant at the 0.01 level (two-tailed)

The relationship between the number of frequency bands making up 95% of tokens in a text, the percentage of tokens from the first three frequency bands and the percentage of tokens from the first frequency band was investigated by Pearson *r*. There was a strong positive correlation between the percentage of tokens from the first frequency bands and the percentage of tokens from the first three frequency bands ($r = 0.709$, $p < 0.01$). This strong positive correlation reveals that both of them increase or decrease in the same direction. When the percentage of tokens from the first three frequency bands is high, it is very likely that the percentage of tokens from the first frequency band is high too and when the percentage of tokens from the first three frequency bands is low, the percentage of tokens from the first frequency band is low as well (see above). The high percentage indicates that a text is easy to read while the low percentage indicates that a text is difficult to read. It is because both of them are based on percentage of tokens from frequency bands especially the first one.

There were moderate negative correlations between the number of frequency bands making up 95% of tokens and the percentage of tokens from the first frequency band (r = -0.453, $p < 0.05$) and between the number of frequency bands and the percentage of tokens from the first three frequency bands (r = -0.683, $p < 0.01$). As mentioned earlier, high numbers of frequency bands indicate texts that are difficult to read and fewer numbers of frequency bands mark texts that are easier to read. This interpretation of the number of frequency bands making 95% of tokens in a text is in the opposite direction from the interpretation of the percentage of tokens from the first three frequency bands and the percentage from the first frequency band (see above). Therefore, the correlations were negative.

From Table 5.9, we can see that the percentage of tokens from the first three frequency bands highly correlated with both the number of frequency bands that make up 95% of tokens in a text and the percentage of tokens from the first frequency band. Additionally, the percentage of tokens from the first three frequency bands also significantly correlated with the results of Flesch readability formula and Flesch-Kincaid readability test (see Table 5.8). Therefore, when we talk about the use of LFP as an indicator of text difficulty, the percentage of tokens from the first three frequency bands seems to be the most suitable indicator of text difficulty.

We have found the most suitable way of using LFP which is the percentage of tokens from the first three frequency bands. However, the validation of LFP as an indicator of text difficulty is still questionable. To give an answer to the question about the validity of this indicator, the validity of LFP as an indicator of text difficulty will be investigated in Section 5.3.

As mentioned in Chapter 2, the indication of text difficulty by LFP is one component of the vocabulary-based readability index. The other two components involve the predictions of reading comprehension. These components employ each of the reader-based variables which are vocabulary size and TSVK in the prediction of reading comprehension.

5.2.2 Prediction of Reading Comprehension at Different Vocabulary Size

This part aims to predict reading comprehension scores based on vocabulary size. In order to do so, the data from the yes/no test (see Appendix A) and the reading comprehension test (see Appendix D) completed by the 102 participants taking part in Phase 3 (the development of the vocabulary-based readability index) were employed. The participants' vocabulary size was estimated from their answers in the yes/no test and their reading comprehension test was marked in order to get their reading comprehension scores on each of the four passages. Then the participants' vocabulary size scores (see Appendix O) and scores on the reading comprehension test with four passages (see Appendix N) were analyzed by simple regression in order to predict reading comprehension scores based on vocabulary size. With the assistance of SPSS 16.0, regression was calculated. The results are presented below.

Table 5.10: Summary of Results of Simple Regression Analysis for the Prediction of Reading Comprehension Scores by Vocabulary Size

Constant/ Independent variable	b	Std. Error	β	t	p-value
Constant	0.138	0.099		1.390	.165
Vocabulary size	5.727	0.169	0.860	33.951	.000

$SE_{est} = \pm 0.862$

R = 0.860 $R^2 = 0.740$ F = 1.153E3 $p = .000$

Note: The dependent variable is reading comprehension scores.

Table 5.10 shows that the vocabulary size and the reading comprehension scores highly correlated at the significance level at $p < 0.01$ (R = 0.860). The R-square (R^2) of the relationship between vocabulary size scores and reading comprehension scores was .740, $p < 0.01$. It gives us some ideas of how well the independent variable (vocabulary size) contributes to the prediction of the scores on the reading comprehension test. This R-square (.740, $p < 0.01$) means that there was a 74 percent possibility that the predictions of reading comprehension scores at different vocabulary size scores tended to be correct with the standard error of estimate at ±0.862. The a value, which is the value of $\hat{y}$ or the dependent variable when the independent variable equals zero, was 0.138 and the b value, which is a coefficient that describes the size of the effect the independent variable is having on the dependent variable, was 5.727. These a and b values were used to form a regression equation for predicting reading comprehension scores. The equation for the prediction of reading comprehension scores ($\hat{y}$) on all of the four reading passages by vocabulary size (X) is: $\hat{y} = 0.138+5.727X$.

The aforementioned regression equation gives us an overview of the use of vocabulary size to predict reading comprehension scores on all of the four reading passages in the reading comprehension test. However, according to Chapter 3, the four texts used in the study had four different LFP. Additionally, from the three-way ANOVA results presented in Table 5.4, the interactions between LFP and vocabulary size scores had some effect on reading comprehension scores so we need to take LFP into the prediction of reading comprehension scores at different vocabulary size scores. The summary of regression results based on each LFP is reported in Table 5.11, Table 5.12, Table 5.13 and Table 5.14 respectively and each of the four equations used for the prediction of reading comprehension scores at different vocabulary size scores are also presented below.

Table 5.11: Summary of Results of Simple Regression Analysis for the Prediction of Reading Comprehension Scores on Reading Passages with 95% of Tokens from the 1st 1000 Most Frequent Word Families (LFP1) by Vocabulary Size

Constant/ Independent variable	b	Std. Error	β	t	p-value
Constant	-0.265	0.272		-0.974	.333
Vocabulary size (LFP1)	6.499	0.373	0.868	17.447	.000
$SE_{est} = \pm 0.685$					
R = 0.868	$R^2 = 0.753$	F = 304.389	$p = .000$		

Note: The dependent variable is reading comprehension scores.

From Table 5.11, it was found that the vocabulary size and the reading comprehension scores highly correlated at the significance level at $p < 0.01$ (R = 0.868). The R-square of the relationship between vocabulary size scores and reading comprehension scores was .753, $p < 0.01$. This means that there was a 75 percent possibility that the predictions of reading comprehension scores at different vocabulary size scores tended to be correct with the standard error of estimate at ±0.685. The a value was -0.265 and the b value was 6.499. These a and b values were used to form a regression equation for predicting reading comprehension scores with 95% of tokens from the first 1000 most frequent word families. The equation for the prediction of reading comprehension scores ($\hat{y}$) by vocabulary size (X) is: $\hat{y} = -0.265+6.499X$.

Only the yes/no test scores ranging from 0 to 1 were substituted in this equation because a participant whose vocabulary size score is lower than 0 seems to have very little vocabulary knowledge and their predicted reading scores are usually negative. From the test specifications of the reading comprehension test presented in

Table 3.5, we can see that it is impossible for the participants to get a negative score. If they cannot answer any question correctly, they will get 0. Therefore, the negative scores on the yes/no test will not be mentioned. If any participants get a yes/no test score which is lower than 0, we assume that they will get 0 on the reading comprehension test. After substituting the equation, the predicted reading comprehension scores are presented in Appendix T.

In Appendix R, there are three columns in the table. Vocabulary size is presented in the first column in terms of numbers of word families. They were calculated by multiplying the scores by 1,000 which is the size of frequency band (see 4.1.1). The second column presents the predicted reading comprehension scores derived from the substitutions. They are actually expressed as decimals. However, according to test specifications (see Table 3.5), the scores on the reading comprehension test are not decimals. The decimals were converted into numbers with the assistance of SPSS 16.0. The round predicted reading comprehension scores are then presented in the third column.

Table 5.12: Summary of Results of Simple Regression Analysis for the Prediction of Reading Comprehension Scores on Reading Passages with 95% of Tokens from the 1st and the 2nd 1000 Most Frequent Word Families (LFP2) by Vocabulary Size

Constant/ Independent variable	**b**	**Std. Error**	**β**	**t**	**p-value**
Constant	-0.614	0.198		-3.107	.002
Vocabulary size (LFP2)	6.893	0.316	0.909	21.845	.000
$SE_{est} = \pm 0.685$					
$R = 0.909$	$R^2 = 0.827$	$F = 477.202$	$p = .000$		

Note: The dependent variable is reading comprehension scores.

Table 5.12 also reveals that the vocabulary size and the reading comprehension scores highly correlated at the significance level at $p < 0.01$ (R = 0.909). The R-square of the relationship between vocabulary size scores and reading comprehension scores was .827, $p < 0.01$. This means that there was an 82 or 83 percent possibility that the predictions of reading comprehension scores at different vocabulary size scores tended to be correct with the standard error of estimate at ±0.685. The a value was -0.614 and the b value was 6.893. These a and b values were used to form a regression equation for predicting reading comprehension scores with 95% of tokens from the first and the second 1000 most frequent word families (LFP2). The equation for the prediction of reading comprehension scores ($\hat{y}$) on texts with 95% of tokens from the first and the second 1000 word families by vocabulary size (X) is: $\hat{y}$ = -0.614+6.893X. The vocabulary size scores ranging from 0 to 1 were substituted into this equation. The predicted reading comprehension scores are shown in Appendix R.

Table 5.13: Summary of Results of Simple Regression Analysis for the Prediction of Reading Comprehension Scores on Reading Passages with 95% of Tokens from the 1st, the 2nd and the 3rd 1000 Most Frequent Word Families (LFP3) by Vocabulary Size

Constant/ Independent variable	b	Std. Error	β	t	p-value
Constant	-0.294	0.185		-1.589	.115
Vocabulary size (LFP3)	6.150	0.348	-0.870	17.678	.000
SE_{est} = ±0.8.20					
R = 0.870	R^2 = 0.758	F = 312.510		p = .000	

Note: The dependent variable is reading comprehension scores.

From Table 5.13, it was also found that the vocabulary size and the reading comprehension scores highly correlated at the significance level at $p < 0.01$ (R = 0.870). The R-square of the relationship between vocabulary size scores and reading comprehension scores was .758, $p < 0.01$. This means that there was a 75 or 76 percent possibility that the predictions of reading comprehension scores at different vocabulary size scores tended to be correct with the standard error of estimate at ±0.820. The a value was -0.294 and the b value was 6.150. These a and b values were used to form a regression equation for predicting reading comprehension scores with 95% of tokens from the first, the second and the third 1000 most frequent word families (LFP3). The equation for the prediction of reading comprehension scores ($\hat{y}$) by vocabulary size (X) is: $\hat{y} = -0.294+6.150X$. The possible vocabulary size scores were substituted into this equation. The predicted reading comprehension scores are given in Appendix R.

Table 5.14: Summary of Results of Simple Regression Analysis for the Prediction of Reading Comprehension Scores on Questions from Reading Passages with 95% of Tokens from the 1st, the 2nd, the 3rd and the 4th 1000 Most Frequent Word Families (LFP4) by Vocabulary Size

Constant/ Independent variable	**b**	**Std. Error**	**β**	**t**	**p-value**
Constant	0.982	0.182		5.401	.000
Vocabulary size (LFP4)	3.706	0.440	0.644	8.419	.000
$SE_{est} = \pm1.019$					
R = 0.644	$R^2 = 0.415$	F = 70.874	$p = .000$		

Note: The dependent variable is reading comprehension scores.

Table 5.14 also reports that the vocabulary size and the reading comprehension scores highly correlated at the significance level at $p < 0.01$ (R = 0.644). The R-square of the relationship between vocabulary size scores and reading comprehension scores was .415, $p < 0.01$. This means that when the equation was used to predict reading comprehension scores on questions of texts with 95% of tokens from the first four word lists, the possibility that the predictions of reading comprehension scores at different vocabulary size scores tended to be correct was only 41 or 42 percent with the standard error of estimate at ±1.019. The a value was 0.982 and the b value was 3.706. These a and b values are used to form a regression equation for predicting reading comprehension scores with 95% of tokens from the first, the second, the third and the fourth 1000 most frequent word families (LFP4). The equation for the prediction of reading comprehension scores ($\hat{y}$) by vocabulary size (X) is: $\hat{y} = 0.982+3.706X$. The vocabulary size scores ranging from 0 to 1 were substituted into this equation. The predicted reading comprehension scores are presented in Appendix R.

When we compared the percent of possibility of the prediction for LFP 4 and the standard error of estimate with the first three, we can see that this one is the lowest with the highest standard error of estimate. In addition, when we look at the a values from Table 5.11, Table 5.12, Table 5.13 and Table 5.14, we can see that the a values in the first three tables were negative while the a value in the last table was positive.

In order to know the LFP of a text, RANGE – computer software for analyzing LFP - should be used (see Chapter 3). The results of RANGE that can be used are in terms of percentage of tokens from each frequency band. Then the number of frequency bands needs to be counted. The number of frequency bands that make up 95% of tokens in the text is used to judge LFP. By way of illustration, look at the LFP presented in terms of percentage of tokens derived from a RANGE analysis of Text 16 in Appendix R in the table below.

Table 5.15: Lexical Frequency Profile of Text 25 in Appendix P

Word List/Frequency Band	Tokens/%
the first 1000	309/88.54
the second 1000	24/6.88
the third 1000	1/0.29
the fourth 1000	5/1.43
the fifth 1000	1/0.29
the sixth 1000	0/0.00
the seventh 1000	5/1.43
the eighth 1000	0/0.00
the ninth 1000	0/0.00
the tenth 1000	0/0.00
the eleventh 1000	0/0.00
the twelfth 1000	0/0.00
the thirteenth 1000	2/0.57
the fourteenth 1000	0/0.00
the fifteenth 1000	0/0.00
the sixteenth 1000	0/0.00
Off-list	2/0.57

From Table 5.15, the first three frequency bands (88.54%, 6.88% and 0.29%) make up 95% of tokens in the text. Hence, we need to use the regression equation for LFP 3 or for reading passages consisting of 95% of tokens from the first, the second and the third 1000 most frequent word families. The four aforementioned equations are presented again below in order to get a picture of how the reading comprehension scores are predicted by vocabulary size when the LFP is different. Each equation is used for each LFP.

LFP1

(reading passages with 95% of tokens from the 1st 1000 most frequent word families) $\hat{y}$ = -0.265+6.499X

LFP2

(reading passages with 95% of tokens from the 1st and the 2nd 1000 most frequent word families) $\hat{y}$ = -0.614+6.893X

LFP3

(reading passages with 95% of tokens from the 1st, the 2nd and the 3rd 1000 most frequent word families)

$\hat{y}$ = -0.294+6.150X

LFP4

(reading passages with 95% of tokens from the 1st, the 2nd, the 3rd and the 4th 1000 most frequent word families)

$\hat{y}$ = 0.982+3.706X

From the aforementioned equations, $\hat{y}$ or reading comprehension scores is the dependent variable that we are trying to predict and vocabulary size score (X) is the independent variable we are using to predict the dependent variable. The numbers -0.265, -0.614, -0.294 and 0.982 were a or the values of $\hat{y}$ or the dependent variable when the independent variable equals 0. The numbers 6.499, 6.893, 6.150 and 3.706 were b or coefficients that describe the size of the effect the independent variable or vocabulary size scores is having on the dependent variable or reading comprehension scores. In order to use the equations to predict the reading comprehension scores, students' or readers' vocabulary size scores should be substituted in the suitable equation.

The predicted reading comprehension scores presented in Appendix R give us a rough idea of how many scores on a reading comprehension test with six questions a student would get when they have a particular vocabulary size. However, it is not practical in terms of text selection because the predicted reading comprehension scores cannot say how easy or difficulty a text is for individual students. Therefore, the predicted reading comprehension scores at different vocabulary size are divided and converted into five bands of text difficulty by range. These bands are: too difficult, difficult, optimal, easy and too easy. In order to calculate range with the assistance of SPSS 16.0, the minimum predicted score was subtracted from the maximum score. The ranges of the four sets (LFP1, LFP2, LFP3 and LFP4) of the predicted scores were 6.50, 6.89, 6.15 and 3.71 respectively. In order to divide the scores into five bands, the each of the ranges was divided by the number of bands (5). Then the results, which were 1.30, 1.38, 1.23 and 0.74, were used to arrange the predicted reading comprehension scores into five equal bands. As a result, predicted reading comprehension scores at different TSVK and five bands of text difficulty are shown below. They are presented based on each of the four LFP.

Table 5.16: Predicted Reading Comprehension Scores at Different Vocabulary Size Scores and Bands of Text Difficulty

	Vocabulary Size Scores	**Predicted Reading Comprehension Scores**		**Bands of Text Difficulty**
		Minimum	Maximum	
LFP1	0-200	-0.26	1.04	too difficult
	201-400	1.05	2.35	difficult
	401-600	2.36	3.65	optimal
	601-800	3.66	4.95	easy
	801-1000	4.96	6.25	too easy
	Vocabulary Size Scores	**Predicted Reading Comprehension Scores**		**Bands of Text Difficulty**
		Minimum	Maximum	
LFP2	0-200	-0.61	0.77	too difficult
	201-400	0.78	2.15	difficult
	401-600	2.16	3.53	optimal
	601-800	3.54	4.91	easy
	801-1000	4.92	6.26	too easy

Table 5.16 (cont.): Predicted Reading Comprehension Scores at Different Vocabulary Size Scores and Bands of Text Difficulty

	Vocabulary Size Scores	Predicted Reading Comprehension Scores		Bands of Text Difficulty
		Minimum	Maximum	
LFP3	0-200	-0.29	0.94	too difficult
	201-400	0.95	2.17	difficult
	401-600	2.18	3.40	optimal
	601-800	3.41	4.63	easy
	801-1000	4.64	5.86	too easy
	Vocabulary Size Scores	**Predicted Reading Comprehension Scores**		**Bands of Text Difficulty**
		Minimum	Maximum	
LFP4	0-200	0.98	1.72	too difficult
	201-400	1.73	2.46	difficult
	401-600	2.47	3.20	optimal
	601-800	3.21	3.94	easy
	801-1000	3.95	4.68	too easy

From Table 5.16, we can see that the predictions of reading comprehension scores on questions from reading passages of four different LFP and the predicted difficulty were similar. If a student gets a vocabulary size score between 401-600 on a frequency band (the 1st 1000, the 2nd 1000, the 3rd 1000 and the 4th 1000 most frequent word families), any texts consisting of 95% of tokens from the frequency band(s) are predicted to be optimal for him/her. For example, if a student gets a vocabulary size score of 500 or 0.50 from the first frequency band consisting of the 1st 1000 most frequent word families, texts with 95% of tokens from the 1st 1000 most frequent word families is predicted to be optimal for him or her.

5.2.3 Prediction of Reading Comprehension at Different Text-Specific Vocabulary Knowledge

This part aims to predict reading comprehension scores based on TSVK. To be able to do the prediction, the 102 university students were asked to 1) complete the reading comprehension test with four passages (see Appendix D); 2) circle unknown words in the self-report on unknown words in the reading comprehension test with four passages (see Appendix D) and 3) complete the translation test for testing TSVK (see Appendix F). Their answers on the reading comprehension test were marked. The TSVK was calculated by Laufer's (1989) two stages of calculation. The participants' TSVK (see Appendix S) and reading comprehension scores (see Appendix N) were analyzed by regression so as to predict reading comprehension scores (dependent variable) based on TSVK (independent variable). The regression was calculated with the assistance of SPSS 16.0. The summary of regression results are presented below.

Table 5.17: Summary of Results of Simple Regression Analysis for the Prediction of Reading Comprehension Scores by TSVK

Constant/ Independent variable	b	Std. Error	β	t	p-value
Constant	-57.999	1.860		-31.175	.000
TSVK	0.635	0.019	0.853	32.888	.000
$SE_{est} = \pm 0.881$ R = 0.853 $R^2 = 0.727$ F = 1.082E3 $p = .000$					

Note: The dependent variable is reading comprehension scores.

According to Table 5.17, it was found that the TSVK and the reading comprehension scores highly correlated at the significance level at $p < 0.01$ (R = 0.853). The R-square of the relationship between TSVK and reading comprehension scores was .727, $p < 0.01$. This means that there was a 72 or 73 percent possibility that the predictions of reading comprehension scores at different vocabulary size scores tended to be correct with the standard error of estimate at ±0.881. The a value was -57.999 and the b value was 0.635. These a and b values were used to form a regression equation for predicting reading comprehension scores. The equation for the prediction of reading comprehension scores ($\hat{y}$) by TSVK (X) is: $\hat{y}$ = -57.999+0.635X.

For the prediction of reading comprehension scores at different TSVK, LFP was not involved because according to the results of three-way ANOVA in Table 5.4, there was no interaction between LFP and TSVK. Therefore, the equation ($\hat{y}$ = -57.999+0.635X) was the only equation for predicting reading comprehension scores at different TSVK. In order to predict reading comprehension scores, the equation was substituted by possible TSVK. The predicted reading scores at round TSVK are presented below.

Table 5.18: Sample Predicted Reading Comprehension Scores at Round TSVK

TSVK (%)	Predicted Reading Comprehension Scores	TSVK (%)	Predicted Reading Comprehension Scores
0	-58.00	55	-23.07
5	-54.82	60	-19.90
10	-51.65	65	-16.72
15	-48.47	70	-13.55
20	-45.30	75	-10.37
25	-42.12	80	-7.20
30	-38.95	85	-4.02
35	-35.77	90	-0.85
40	-32.60	95	2.33
45	-29.42	100	5.50
50	-26.25		

As mentioned in the prediction of reading comprehension at different vocabulary size scores, the predicted reading comprehension scores only give us a rough idea of how many scores on a reading comprehension test with six questions a student would get when they have a particular TSVK. However, it does not seem to help in terms of text selection because the predicted reading comprehension scores do not tell teachers and students who would like to select a text whether the text is suitable for their students or for themselves or not. Therefore, the predicted reading comprehension scores at different TSVK are divided into five bands of text difficulty by range in the same way as the prediction of reading comprehension at different vocabulary size. These bands are: too difficult, difficult, optimal, easy and too easy.

Since the a value (-57.999), which is the value of $\hat{y}$ or the predicted reading comprehension scores when the TSVK equals 0, in Table 5.18 was very low, most of the predicted reading scores at different TSVK ranging from 0-91% are negative (see Table 5.18). As mentioned earlier, the negative scores on the reading comprehension test are not possible. Students will be predicted to get a score between 0-6 when their TSVK ranges from 92%-100%. Hence, the predicted reading comprehension scores which are lower than 92% will be ignored. Only the reading comprehension scores predicted at different TSVK ranging from 92%-100% are arranged in the five bands of text difficulty (see Appendix T).

The range was 5.08. Then the range (5.08) was divided by the number of bands (5) and the result was 1.02. This number (1.02) was used to arrange the predicted reading comprehension scores into five equal bands. As a result, predicted reading comprehension scores at different TSVK and five bands of text difficulty are presented below.

Table 5.19: Predicted Reading Comprehension Scores at Different TSVK and Bands of Text Difficulty

TSVK (%)	**Predicted Reading Comprehension Scores**		**Bands of Text Difficulty**
	Minimum	Maximum	
92.00-93.61	0.42	1.44	too difficult
93.62-95.21	1.45	2.46	difficult
95.22-96.82	2.47	3.48	optimal
96.83-98.43	3.49	4.50	easy
98.44-100.00	4.51	5.50	too easy

Table 5.19 shows that a text is predicted to be optimal for a student when the student knows 95%-96.82% of tokens in the text. This is similar to Laufer (1989) reporting that students who know 95% of tokens in a text can comprehend that text. However, Table 5.19 showed that a text is said to be optimal when the predicted reading scores are between 2.47 to 3.48 which are 41% to 58% while Laufer and Sim found that 65% to 70% was the minimum score where readers would have enough language knowledge to comprehend a text.

5.3 Validation of Vocabulary-based Readability Index

The primary aim of this section is to investigate validity of the vocabulary-based readability index in order to answer the research question “To what extent would a purely vocabulary-based readability index be valid?”. In order to answer the question, the terms “validation” and “validity” should be taken into considerations. The term “validation” in language testing is generally defined as “a process of gathering evidence to support the claim that a test measures certain abilities or attributes in certain contexts for certain purposes” (Douglas, 2000). Correspondingly, when the term is applied to a validation of the vocabulary-based readability, it is defined based on the definition of readability in Chapter 2 so the term "validation" in the present study refers to the process of collecting evidence to support the claim that the index measures the level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts.

The term "validity" is defined in language testing as "the degree to which a test measures what it is supposed to measure, or can be used successfully for the purposes for which it is intended" (Richards and Schmidt, 2002). Similarly, in the validation of the vocabulary-based readability index, it refers to the degree to which the index measures the level or degree of the ease or difficulty of texts for Thai university students to comprehend texts.

Different authors such as Hughes (1989; 2003); Alderson, Clapham and Wall (1995); Johnson (2001); and Brown (2005) divide validity into different types. According to Hughes (1989; 2003), there are 4 types of validity: content validity, predictive validity, face validity and criterion-related validity. Hughes also suggests that concurrent validity and construct validity are subcategories of criterion-related validity. Similarly, Johnson (2001) divides validity in the same way as Hughes but he calls some of them in a different way. He states that there are 5 types of validity: content validity, face validity, construct validity, empirical or criterion-related validity and predictive validity. His definition of empirical or criterion-related validity is the same as the definition of concurrent validity given by Hughes. In the same way, Brown (2005) divides validity into three main types: content, construct and criterion-related validity. He also suggests that concurrent and predictive validity are variations of criterion-related validity but he does not include face validity as a type of validity. Correspondingly Alderson et al. (1995) present some similar types of validity but they add some more types and use a different way to present the types of validity. They divide validity into 3 main types: internal, external and construct validity. Internal validity consists of face, content and response validity. External validity is like Hughes' (1989; 2003) and Brown's (2005) criterion-related validity. It has two subcategories which are concurrent and predictive validity.

From Hughes (1989; 2003); Alderson et al. (1995); Johnson (2001); and Brown (2005), we can see that all of them present three similar types of validity which are content validity, criterion-related validity including concurrent and predictive validity and construct validity. However, since it is difficult to look for experts in the field of readability so it will be better if we investigate face validity rather than content validity. Moreover, the construct validity will not be investigated in this chapter. It will be discussed in Chapter 6. This section is then divided into three main sections:

1) Face validity in the present study refers to the degree to which the vocabulary-based readability index appears to predict level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts. It was investigated by telling 3 university lecturers and 3 students the purpose of the index, which is to predict the level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts, and how to use of LFP, vocabulary size and TSVK in the predictions of text difficulty and how easy or difficult a text is for individual students. They were asked to rate the extent that they agreed that the vocabulary-based readability index looks good as a predictor of level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts on the

5-point scale. The scale consists of 1 (extremely disagree), 2 (disagree), 3 (unsure), 4 (agree) and 5 (extremely agree).

2) Concurrent validity refers to the extent to which a test correlates with some other test that is aimed at measuring the same skill, or with some other comparable measure of the skill being tested (Richards and Schmidt, 2002). In order to investigate concurrent validity of LFP as a predictor of text difficulty, Spearman's rank correlation or Spearman's rho was employed so as to investigate the relationship between the rankings of text difficulty ranked by LFP, lecturers and students. Moreover, Pearson *r* was also used to investigate the relationship between the percentage of tokens from the first three frequency bands, the results from Flesch readability formula and Flesch-Kincaid readability test and the 45 university students' reading scores on the reading comprehension test with four passages.

3) Predictive validity is defined in this study as the degree to which the index accurately predicts level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts. According to 5.2.2 and 5.2.3, the regression equations were used to predict reading comprehension scores at different vocabulary size and at different TSVK. After predicting reading comprehension scores, the predicted scores were arranged into five bands of text difficulty in order to predict how easy or difficult a text is for individual students to comprehend the text so we need to validate as two types of predictions which are: 3.1) the extent that the index accurately predicts reading comprehension scores and 3.2) the extent that the index accurately predicts how easy or difficult a text is for individual students.

The face, concurrent and predictive validity were investigated. The findings are presented in 5.3.1, 5.3.2 and 5.3.3 respectively.

5.3.1 Investigation of Face Validity of the Vocabulary-based Readability Index

Before investigating concurrent and predictive validity, it is interesting to investigate face validity in order to learn university lecturers' and students' opinions whether the index looks good as a measure of the level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts. Three teachers and three students were told the purpose of the index and shown how to predict text difficulty and reading comprehension. Then they were asked to rate the extent that they agreed that the index looks good as an indicator of text difficulty and predictor of reading comprehension. There were 5 points on the rating scale: 1 (extremely disagree), 2 (disagree), 3 (unsure), 4 (agree) and 5 (extremely agree). The results are presented based on the three components of the index, the indication of text difficulty by LFP, the prediction of reading comprehension at different vocabulary size and at different TSVK.

Table 5.20: Respondents' Opinion about the Face Validity of the Indication of Text Difficulty by LFP

Respondent	Extremely Disagree	Disagree	Unsure	Agree	Extremely Agree
L1				✓	
L2				✓	
L3			✓		
S1			✓		
S2			✓		
S3			✓		

Note: L stands for lecturer
S stands for student

From Table 5.20, two out of three of the teachers (L1 and L2) agreed that LFP looks good as an indicator of text difficulty. The other teacher and all of the students were not so sure that LFP can be used to indicate text difficulty.

Table 5.21: Respondents' Opinion about the Face Validity of the Prediction of Reading Comprehension at Different Vocabulary Size

Respondent	Extremely Disagree	Disagree	Unsure	Agree	Extremely Agree
L1			✓		
L2			✓		
L3			✓		
S1				✓	
S2				✓	
S3				✓	

Note: L stands for lecturer
S stands for student

The results from Table 5.21 are quite different from the ones in Table 5.20. All of the lecturers were not so sure that following the instructions of how to predict reading comprehension and predicting reading comprehension in terms of bands of text difficulty can help them select appropriate texts. For the students' opinion, all of them agreed that vocabulary size and bands of text difficulty can be used to predict reading comprehension.

Table 5.22: Respondents' Opinion about the Face Validity of the Prediction of Reading Comprehension at Different TSVK

Respondent	Extremely Disagree	Disagree	Unsure	Agree	Extremely Agree
L1			✓		
L2			✓		
L3			✓		
S1				✓	
S2				✓	
S3				✓	

Note: L stands for lecturer
S stands for student

All of the respondents' opinion about the prediction of reading comprehension at different TSVK was in the same way as the results in Table 5.21. The lecturers were unsure that TSVK can be used to predict reading comprehension while the students agreed that TSVK and the bands of text difficulty can give them some ideas for text selection.

5.3.2 Investigation of Concurrent Validity of Lexical Frequency Profile as an Indicator of Text Difficulty

According to Chapter 1, LFP refers to proportion of low and high frequency words in a text. It presents us percentage of tokens in a text from each word list. Texts with many high frequency words tend to be easier than texts with many low frequency words. It is used to predict text difficulty by calculating the percentage of tokens from the first three frequency bands (see 5.2.1).

In order to investigate whether we can use LFP to indicate text difficulty or to predict the level or degree of the ease or difficulty of texts, we need to investigate the relationship between the results of LFP and: 1) rankings of text difficulty of four reading passages ranked by 45 university students (see Appendix U) in the questionnaire for investigating concurrent validity of LFP as an indicator of text difficulty (see Appendix J); 2) five experienced lecturers' rankings of the four reading passages in terms of text difficulty (see Appendix V) in the questionnaire for investigating concurrent validity of LFP as an indicator of text difficulty (see Appendix J) and 3) results or readability scores from Flesch readability formula and Flesch-Kincaid readability test (see Appendix W). When we look at these four comparisons, we can see that two types of data which are ordinal and interval are involved. The participants' and teachers' rankings are ordinal. When the data are ordinal, Spearman's rho needs to be used to investigate the correlation. The participants' scores on the reading comprehension test and the results of Flesch readability formula ranging from 0-100 and Flesch-Kincaid readability test ranging from 0-12 are interval. When the data are interval, Pearson *r* needs to be used.

Both Spearman's rho and Pearson *r* are measures of correlation between two variables (Wikipedia, www, 2010). They vary between -1 and +1. Any correlation near +1 indicates the very high level of agreement among the indicators of text difficulty while a correlation of -1 means that there is a perfect negative relationship between variables and a correlation of 0 means that there is no relationship between the two variables. The results of Spearman's Rho and Pearson *r* are presented in Table 5.23 and 5.24 respectively.

Table 5.23: Rank Correlations between Text Difficulty Indicated by LFP and Other Rankings of Text Difficulty

	Students' rankings	Teachers' rankings
LFP	ρ = .822** p = 0.000	ρ = 0.840** p = 0.000

** Correlation is significant at the 0.01 level (two-tailed)

Table 5.23 shows that the ranking of difficulty indicated by LFP highly correlated with the rankings obtained from the participants' and lecturers' rankings (p < 0.01). The correlation coefficients were .822 and .840 respectively. These results were higher than Lunzer and Gardner (1979) who conduct readability research for the Schools Council Effective Use of Readability project. In their research, they investigated validity of eight formulas by studying the relationship between the results from these formulas with the teachers' judgments and that of Dale and Chall (1948) had the highest correlation (r = .77).

For the investigation of the relationship between the results of LFP or the percentage of tokens from the first three frequency bands and the participants' performance on the reading comprehension test and results from Flesch readability formula and Flesch-Kincaid readability test, Pearson *r* was employed. The results are presented below.

Table 5.24: Pearson's Correlation Coefficients between the Percentage of Tokens from the First Three Frequency Bands (LFP) from the Four Reading Passages and Results of Two Readability Formulas

	Flesch Readability Formula	**Flesch-Kincaid Readability Test**
Percentage of Tokens from the First Three Frequency Bands	$r = 0.873$** $p = 0.000$	$r = -0.992$** $p = 0.000$

** Correlation is significant at the 0.01 level (two-tailed)

As mentioned earlier, Flesch readability formula and Flesch-Kincaid readability test yield readability results in opposite directions. For Flesch readability formula, a high result or reading ease of a text indicates that the text is easy while the high American grade level presented as the result of Flesch-Kincaid readability test indicate that the text is difficult to read. Hence, the correlation coefficients gained from the correlation between the percentage of tokens from the first three frequency bands and the results of the two readability formulas, Flesch readability formula and Flesch-Kincaid readability test are undoubtedly opposite. The significant correlation between the percentage of tokens from the first frequency bands and the results of Flesch readability formula was positive and strong ($r = 0.873$, $p < 0.01$) because when the percentage of tokens from the first three frequency bands of a text is high, that text is indicated as easy to read. Similarly, the higher reading ease score, which is the results of Flesch readability formula, marks the text as easier to read. Conversely, there was a strong and negative correlation between the percentage of tokens from the first three frequency bands and the American grade level predicted by Flesch-Kincaid readability test (r = -0.992, p < 0.01). This is because the higher the predicted American grade level, the more difficult the text is while the higher the percentage of tokens, the easier the text is for readers. It can be concluded from the correlation coefficients from Table 5.24 that the extent to which the index correlates with the two traditional readability formulas was high and it is likely that LFP as an indicator of text difficulty has concurrent validity with the two readability formulas.

In order to investigate the possibility that LFP would be valid as an indicator of text difficulty, the correlation coefficients were converted to R-square by squaring the correlation coefficients or r values. The R-squares of the correlations between the percentage of tokens from the first three frequency bands which is the results of LFP and Flesch readability formula and Flesch-Kincaid readability test were 0.76 and 0.98 respectively. This means that there was a 76 percent possibility that the extent that the use of LFP to indicate text difficulty was as concurrently valid as Flesch readability formula and there was a 98 percent possibility that the indication of text difficulty by LFP was as concurrently valid as Flesch-Kincaid readability test.

5.3.3 Investigation of Predictive Validity of Predictions of Reading Comprehension

As mentioned in Chapter 2 that there are two means of predicting how easy or difficult a text is for individual students: 1) prediction of reading comprehension scores at different vocabulary size and 2) prediction of reading comprehension scores at different TSVK. These predictions were done by regression equations. The equations were substituted so as to show a rough idea of how reading scores are predicted in two different ways. The predicted reading scores were arranged into five bands of text difficulty: too difficult, difficult, optimal, easy and too easy. This means that the index predict both reading comprehension and bands of text difficulty. Then it is necessary to investigate predictive validity – the degree to which the index accurately predicts reading comprehension scores and how easy or difficult a text is for individual students to comprehend it.

In order to investigate the predictive validity, the predicted reading comprehension scores were compared with comprehension test scores and subjective judgments (Harrison, 1980). To be able to do so, the comparison of participants' actual test scores and the predicted reading comprehension scores were investigated by Pearson *r* in order to investigate the extent that the index accurately predicts reading comprehension scores. Additionally, the comparison of subjective judgments on 5-point rating scales ranging from 1 (too difficult), 2 (difficult), 3 (optimal), 4 (easy) and 5 (too easy) and the predicted difficulty in terms of how easy or difficult a text is for individual students to comprehend the text was investigated in order to study the extent that the predicted difficulty are accurate.

In order to investigate predictive validity of the predicted reading comprehension scores, 80 university students were asked to complete the yes/no test, the self-report on unknown words for testing TSVK of reading passages in the reading comprehension test with five passages from TOEIC tests, the translation test for testing TSVK in the reading passages in the reading comprehension test with five passages from TOEIC tests and the reading comprehension test with five passages from TOEIC tests.

According to Chapter 3, the yes/no test were scored by *h - f*. Also the reported unknown words in the self-report on unknown words were compared with the answers in the translation test in order to calculate TSVK. When we got their scores on the yes/no test and the TSVK, these scores were substituted in the equations for predicting reading comprehension scores at different vocabulary size and at different TSVK respectively. The results from the equations were compared with their actual scores or scores on the reading test with five passages from TOEIC tests by Pearson *r*. The 80 participants' reading scores on the reading comprehension test are given in Appendix X and the Pearson product-moment correlation coefficients are shown below:

Table 5.25: Correlation Coefficients among Participants' Reading Scores and Reading Scores Predicted by Vocabulary Size Scores and TSVK (N = 400)

	Predicted reading scores at different vocabulary size scores	**Predicted reading scores at different text-specific vocabulary knowledge**
Participants' reading scores	$r = 0.784$** $p = .000$	$r = 0.236$** $p = .000$

** Correlation is significant at the 0.01 level (two-tailed)

Table 5.25 shows that the predictions of reading comprehension scores predicted by vocabulary size scores and TSVK significantly correlated with the participants' actual reading scores at $p < 0.01$. This implies that when predicted reading scores at different vocabulary size and at different TSVK were high, the participants' actual reading scores were also high and when the predicted scores were low, the participants' reading scores were also low.

Furthermore, when we squared the coefficient of the correlation between the predicted reading scores at different vocabulary size scores (r = 0.784, p < .01) in Table 5.25 we can see that there is a 61% or 62% of possibility that the vocabulary size could predict the participants' actual reading comprehension scores. However, when the correlation coefficient between the predicted reading scores at different TSVK (r = 0.236, p < .01) was squared, the possibility that TSVK can predict reading comprehension scores was much lower. There was only 8.496% of possibility. Therefore, it is very likely that the extent that vocabulary size scores accurately predict reading comprehension scores seems to be greater than TSVK.

In order to investigate the predictive validity of the predicted difficulty or how easy or difficult a text is for individual students to comprehend it, the 30 university students were asked to complete the yes/no test, the self-report on unknown words for testing TSVK in the reading passages in the reading comprehension test with four passages, the translation test for testing TSVK in the reading passages in the reading comprehension test with four passages and rate each of the four reading passages in the questionnaire for investigating predictive validity of the prediction of reading comprehension in terms of text difficulty on 5-point scales ranging from 1 (too difficult) to 5 (too easy). The relationship between the students' ratings of text difficult (see Appendix Y) in the questionnaire for investigating predictive validity and the predicted levels of text difficulty by vocabulary size (see Appendix Z) and TSVK (see Appendix AA) were investigated by Pearson r. The results are presented below.

Table 5.26: Correlation Coefficients among Participants' Rating of Text Difficulty and Predicted Difficulty (N = 120)

	Predicted text difficulty by vocabulary size scores	**Predicted reading text difficulty by TSVK**
Students' rating of text difficulty	0.804** *p* = .000	0.676** *p* = .000

** Correlation is significant at the 0.01 level (two-tailed)

Table 5.26 shows the positive correlation coefficients between the predictions of how easy or difficult a text is for individual students to comprehend by vocabulary size scores and TSVK and the participants' ratings of text difficulty were positive. This means that the prediction of how easy or difficult a text is and the participants' ratings of text difficulty rated the text difficulty in the same direction. Additionally, both of the correlation coefficients significantly correlated with the participants' ratings of text difficulty at $p < 0.01$. When we squared the coefficient of the correlation between the predicted difficulty of text for each student to comprehend it ($r = 0.804$, $p < .01$), we can see that there is a 64% or 65% of possibility that the vocabulary size could predict how easy or difficult a text is for individual students to comprehend. This percent of possibility seems to be similar to the possibility that the vocabulary size could predict the participants' actual reading comprehension scores. Moreover, when we look at the correlation coefficient between the predicted difficulty for each student to comprehend the text and the participants' ratings of text difficulty ($r = 0.676$, $p < .01$), we can see that it was much higher than the correlation coefficient between the predicted reading comprehension scores at different TSVK and the participants' reading scores ($r = 0.236$, $p < .01$). When we squared this correlation coefficient ($r = 0.676$, $p < .01$), the possibility that TSVK can predict how easy or difficult a text is for individual students to comprehend was 45% or 46%.

Therefore, it can be concluded that there seems to be some possibility that vocabulary size and TSVK can accurately predict how easy or difficult a text is for individual students to comprehend.

5.4 Summary

To sum up, this chapter showed the results of the present study. The statistical analyses which are three-way ANOVA, regression, Pearson *r* and Spearman rho were used to analyze the data in order to devise the vocabulary-based readability index and investigate the extent that the index would be valid in terms of face, concurrent and predictive validity. The findings revealed that it is very likely that LFP can be used as an indicator of text difficulty and vocabulary size seems to be a promising predictor of reading comprehension scores while the prediction of reading comprehension scores at different TSVK seems to be a less hopeful predictor.

CHAPTER 6

DISCUSSION AND IMPLICATIONS

Based on Chapter 5, three components of the vocabulary-based readability index which were 1) the indication of text difficulty by LFP; 2) the prediction of reading comprehension at different vocabulary size and 3) the prediction of reading comprehension at different TSVK were devised. This chapter then presents the discussion of research findings relating to these three components of vocabulary-based readability index in terms of face validity, concurrent validity, predictive validity and construct validity. It also provides implications for teachers' and students' text selection. The chapter ends with methodological implications and suggestions for further studies.

6.1 Discussion

According to Chapter 1, this study aims to devise a readability index called vocabulary-based readability index and investigate its validity. The index was devised in the hope that language teachers and Thai university students might be able to predict level or degree of the ease or difficulty of a text for individual students to comprehend the texts in order that they can select texts which are suitable for reading comprehension of individual university students. In order to do so, both text-based variable (LFP) and the reader-based variables (vocabulary size and TSVK) were combined in the development of the index.

Before the index was devised, the effects of these variables and their interactions were investigated whether they had any effect on the reading comprehension scores. The results of three-way ANOVA from Table 5.4 revealed that there were significant differences between mean reading comprehension scores when LFP, vocabulary size and TSVK were varied. Additionally, the interaction between LFP and vocabulary size seemed to have some effects on the reading comprehension scores. On the other hand, the results of three-way ANOVA did not show that the interactions between LFP and TSVK; TSVK and vocabulary size; and LFP, vocabulary size and TSVK had any effect on the reading comprehension scores. This is because LFP and vocabulary size were both based on the four word lists from BNC while TSVK involves percentage of known words in a text. Although it is one of the many factors affecting reading comprehension (Hirsh and Nation, 1992), it has nothing to do with either word frequency or word lists from the BNC.

The results of the Scheffé test on differences of pairs of mean reading scores arranged by LFP (see Table 5.5) showed that the participants' reading scores tend to significantly decrease when the number of high frequency words decrease and vice versa. These results made the use of LFP to indicate text difficulty promising because they supported the claim that the percentage of high frequency words in a text can predict text difficulty (Laufer, 1997). Furthermore, the results of the Scheffé test on differences of pairs of mean reading scores arranged by vocabulary size (see Table 5.6) confirmed that knowing more words would assure better comprehension of reading passages (Horst, Cobb and Meara, 1998).

According to Chapter 5, the vocabulary-based readability index was devised and validated based on the definition of readability which is the level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts. There were two main parts of the index: 1) the prediction of text difficulty or the level or degree of the ease or difficulty of texts and 2) the prediction of reading comprehension in terms of how easy or difficult a text is for each student to comprehend the text. These two parts are called predicting and measuring readability respectively (Klare, 1974-1975). These two parts were validated in order to answer the research question "To what extent would a purely vocabulary-based readability index be valid?" In the validation process in Chapter 5, three types of validity which are face validity, concurrent validity and predictive validity were investigated. The validation will be discussed below.

6.1.1 Face Validity of the Three Components of Vocabulary-based Readability Index

The investigation of face validity in 5.3.1 revealed that the lecturers and the students tended to have different ideas about what should be used to select appropriate texts. For the use of LFP, two of the teachers agreed that it looked good as an indicator of text difficulty because they tended to realize that proportion of low and high frequency words seems to have some effects on text difficulty. On the other hand, all of the students tended to believe that number of high frequency words did not mean anything to them because they thought that they did not have much vocabulary knowledge. Although words in a text frequently occur in everyday life, it does not mean that the words must be known to them.

For the prediction of reading comprehension at different vocabulary size, all of the teachers felt unsure about it because it did not seem practical. For students, all of them tended to believe that their vocabulary size had some effects on reading comprehension. However, they seemed to be hesitant to say extremely agree because they did not think that only 30 words randomly selected from 1000 word families could represent their vocabulary knowledge in the entire word list. Two of them insisted that they knew most of the words in level 1 in the yes/no test but there were many words in the first word list that they did not know.

Similarly, all of the lecturers were unsure whether TSVK could be used to predict reading comprehension because although it is a good idea to select appropriate texts for students, it is believed in language teaching that the texts used for instruction should have not been mastered by the students (Lazar, 2004). Furthermore, the teachers' feeling of unsure might be due to the detailed process of analyzing TSVK or the lack of practicality. For students, they agreed that TSVK can be used to predict reading comprehension because when they know more words in a text, they tend to be able to read the text better.

6.1.2 Concurrent Validity of the Indication of Text Difficulty by Lexical Frequency Profile

LFP is the only text-based variable in this study. The analysis of LFP involves counting or calculating percentage of high frequency words in a passage in order to provide an index of probable difficulty for readers. It is a predictive device in the sense that no actual participation by readers is needed (Klare, 1974-1975).

According to the investigation of concurrent validity of LFP and an indicator of text difficulty (see 5.3.2), the correlation coefficients from Spearman's rho and Pearson *r* revealed that LFP indicates text difficulty in the same way as Flesch readability formula, Flesch-Kincaid readability test, lecturers as well as students. These similarities may be caused by some aspects of methodology that the analysis of LFP and the readability formulas have in common. When we look at how the analysis of LFP and the two readability formulas indicate text difficulty, there are two explicit similarities among them: 1) all of them use the counts of difficult words (Davey, 1988) and 2) the counting unit is tokens - any occurrence of a word form in the text, regardless of whether it is occurring for the 1^{st} or the n^{th} time (Bauer and Nation, 1993). These similarities might lead to the high correlation coefficients from Table 5.24.

For the differences, there are four different aspects: elements employed in the prediction of text difficulty, identifications of difficult words, calculations and presentation of results. For the elements employed in the prediction of text difficulty, LFP is the only variable in the prediction of text difficulty while according to Chapter 1, Flesch readability formula and Flesch-Kincaid readability test employ three variables which are number of words, number of sentences and number of syllables. The number of words and the number of sentences are used to estimate text complexity and as mentioned earlier, the number of syllables is used to estimate number of difficult words. It is surprising that LFP does not take any sentence length or text complexity in the prediction but from Table 5.24, the results of LFP correlated quite well with the results of Flesch readability formula and Flesch-Kincaid readability test. This tends to support some of the findings of Davison and Kantor (1982) that sentence length was not the actual feature of texts that make them easy or difficult to understand.

The identifications of difficult words are also different. The analysis of LFP concerns counting numbers of high frequency words from the first three frequency bands and converts them into percentage because according to 2.5.1, it is claimed that the students tend to be familiar with high frequency words so a text with a lot of high frequency words tends to be easy while the one with a lot of low frequency words tends to be more difficult. For example, Passage One consisting of 95% of tokens from the first 1000 most frequent word families tends to be easier than Passage Four with 95% of tokens from the first four frequency bands. On the other hand, Flesch readability formula and Flesch-Kincaid readability test making use of vocabulary difficulty do not focus on frequency of word occurrences. They try to objectively identify vocabulary difficulty by word length in terms of average numbers of syllables instead because it is believed that difficult words tend to be longer than easy words. It is interesting that LFP and the traditional readability formulas use different ways of identifying text difficulty but they tend to yield similar predictions of text difficulty. This implies that there might be some relationship between word frequency and word length.

For the calculation of the traditional formulas, the formulas integrate the investigated variables such as number of words, number of sentences and number of syllables into a regression equation used to predict text difficulty while regression is not used for indicating text difficulty by LFP. For LFP analysis, after analyzing LFP by RANGE, we need to combine numbers of words from the first three frequency bands and convert it to percentage. This reveals that the calculation of LFP seems to be simpler than the calculation of the traditional formulas.

For the presentation of results, Flesch readability formula presents readability results in terms of reading ease scores. It rates a text on a 100-point scale based on the average number of syllables per word and words per sentence. The higher the Flesch reading ease score, the easier it is to understand the text. Similarly, the results of the analysis of LFP which are presented in terms of percentage of tokens from the first three frequency bands range from 0-100. The higher the percentage of tokens, the easier the text is. On the other hand, Flesch-Kincaid Grade Level rates a reading passage on an American grade-school level from 1 to 12. The higher the grade level, the more difficult it is to understand the text. Although Flesch-Kincaid readability test presented the results in the opposite direction of LFP and Flesch readability test, the correlation coefficients in Table 5.24 showed that all of them predict text difficulty in the same way. That is to say, a text predicted to be an easy text to read by LFP tended to be predicted as an easy one by Flesch readability formula and Flesch-Kincaid readability test.

6.1.3 Predictive Validity of the Prediction of Reading Comprehension

According to Chapter 5, when reading comprehension scores and how easy or difficult a text is for individual students to comprehend the text were predicted, it is important to investigate the extent that the predictions were accurate. Hence, predictive validity, which was defined as the degree to which the index accurately predicts reading comprehension scores and how easy or difficult a text is for individual students to comprehend it, was investigated.

When the predictive validity of predictions of reading scores at different vocabulary size were validated in Chapter 5, the high Pearson product-moment correlation coefficient between the predicted reading scores at different vocabulary size and the participants' reading scores suggest that the index is doing an effective job in predicting reading comprehension (Harrison, 1980). Moreover, the correlation coefficient of the relationship between the reading scores predicted by vocabulary size and the participants' reading scores is higher than a correlation of 0.77 in Dale and Chall (1948) which is claimed that this is the highest correlation among that of eight formulas in readability research carried out for the Schools Council Effective Use of Reading Project (Lunzer and Gardner, 1979). Additionally, the correlation coefficient between participants' reading scores and reading scores predicted by vocabulary size scores are also higher than the correlations of .64 and .70 among the revised Flesch formula scores and the 1950 McCall-Crabbs scores and the 1925 McCall-Crabbs scores (Klare, 1974-1975).

Although the *p*-value at $p < .01$ showed a significant correlation between participants' reading scores and the predicted reading scores at different TSVK, the Pearson product-moment correlation coefficient was low. This might be because the translation test used for testing TSVK was not properly constructed. According to Chapter 3, all of the participants in all of the four phases majored in science-oriented fields but the target words in the translation tests were selected by asking the undergraduates majoring in non-science-oriented fields at Silpakorn University to circle unknown words.

6.1.4 Construct Validity of the Vocabulary-based Readability Index

Construct validity refers to the extent that a test measures the construct it is supposed to measure (Brown, 1996). There are several ways to investigate construct validity such as content analysis, factor analysis, ANOVA (Brown, 2000). However, it was not investigated in this study because some threats to construct validity have been found and these threats tend to destroy the construct validity of the vocabulary-based readability index.

As mentioned earlier, there were three components of the vocabulary-based readability index: 1) the indication of text difficulty LFP; 2) the prediction of reading comprehension at different vocabulary size and 3) the prediction of reading comprehension at different TSVK. There components involve four variables. These variables were LFP, vocabulary size, TSVK and reading comprehension. These variables are sub-constructs of the vocabulary-based readability index. Hence, in order to talk about the construct validity of the vocabulary-based readability index, it is important to consider how to measure each of them and whether each measure really measures the construct it claims to measure.

The data on these four variables were obtained from different measures. LFP was analyzed by RANGE based on BNC. Vocabulary size was tested by yes/no test. TSVK was tested by a self-report on unknown words as well as a translation test and then analyzed by Laufer's (1989) formula or calculation of TSVK. Reading comprehension was tested by a reading comprehension test. These measures will be discussed in terms of construct validity below.

RANGE has been designed and developed by Nation and Heatley (computer software, 2002). It compares an electronic text against vocabulary lists called baseword lists to see what words in the text are and are not in the lists and to see what percentage of the words in the text are covered by the lists (Nation, 2005). It was used in this study to analyze one of the constructs which is LFP or proportion of low and high frequency words in everyday life. Since it is computer software, there is no doubt that RANGE really compares a text with vocabulary lists or what it is instructed or set to do. However, it is questionable whether the words categorized as high frequency words in the word lists are really high frequency words in the contexts of Thai university students. The word lists are based on spoken section of BNC. It is not certain that high frequency words in spoken English are high frequency words that the students normally find in reading. Moreover, the samples from the BNC spoken section including a large amount of unscripted informal conversation, recorded by volunteers selected from different age, region and social classes in a demographically balanced way, together with spoken language collected in all kinds of different contexts, ranging from formal business or government meetings to radio shows and phone-ins" (University of Oxford, www, 2005) are generally language used in British contexts. The words frequently used in those contexts may not be used in the students' contexts. For example, when we talk about a person whose job is to take care of a large building, such as a school, and who deals with the cleaning, repairs, etc., most students seem to be more familiar with the word "janitor" which is American and Scottish English than caretaker or porter which is British English because the word "janitor" is shown in both written and spoken forms on several famous movies such as Good Will Hunting, Catch Me If You Can and The Janitor. Surprisingly, this word is not in the 16 word lists of the BNC spoken section. This means that the word is considered as a low frequent word in everyday life according to spoken section of the BNC. As a result, it is uncertain that the analysis of LFP based on the spoken section of BNC really presents proportion of low and high frequency words in the students' contexts and the use of LFP to indicate text difficulty based on the percentage of the first three frequency bands might be in doubt.

For vocabulary size, since the present study focuses on reading, only receptive vocabulary size was investigated. It is then defined as quantity of reader's general vocabulary knowledge in terms of recognition of written form of words and their meaning. It was tested by a yes/no test. According to Mochida and Harrington (2006) and the results of the pilot work in Chapter 4, it is very likely to say that the yes/no test tended to have high reliability and concurrent validity. However, the construct validity is still uncertain. The only thing that seems to show that the yes/no test tended to have construct validity is the instructions. From the instructions (see Appendix A), the yes/no test requires the participants to read the words and circle the words that they know well enough to say what they mean. In other words, the students are supposed to circle the words that they can recognize and know the meanings. However, the yes/no test is like self-assessment of vocabulary knowledge requiring the students to judge their own knowledge (Oscarson, 1997) so it is difficult to say whether the students really know the words that they circled. Fortunately, pseudowords were included in the hope that they would help distinguish the students who are really familiar or know the target words from the students who circle the words by chance. However, there are no clear guidelines for constructing pseudowords (Eyckmans, www, 2004) and the extent to which pseudowords should differ from real words is still unclear (Abels, 1994 quoted in Eyckmans, www, 2004). Hence, the pseudowords that have similar spelling to the target words might distract the students especially the ones suffering from decoding problems or dyslexia. The other thing that might weaken the construct validity of the yes/no test is that some words have more than one meaning but the yes/no test does not permit the testing of multiple meanings of the target words (Abels, 1994 quoted in Eyckmans, www, 2004).

For TSVK, it was investigated by a self-report on unknown words, a translation test and analyzed by Laufer's (1989) formula. In order to see whether TSVK has construct validity, it is important to look at the construct of the self-report on unknown words, the translation test and the formula and discuss each of them in detail. The self-report on unknown words aims to measure the participants or students unknown words. From the view of face validity, it is so obvious that the self-report on unknown words measures unknown words because we ask the students to circle words in order to report the words that they cannot give the meaning. However, it is doubtful whether the students tell the truth. The report of unknown words is purely based on the students' decisions and their decisions might be affected by several factors. For instance, they might not be able to give the meaning to some familiar words but they want to pretend to be good by ignoring such words; the students did not have enough time to look through the text and circle the unknown words; the students might be too lazy to look through every word in the text and circle all of the unknown words; some of them might be worried that the report of too many unknown words might affect their reading scores and the students might report too many unknown words because they are afraid that their teacher would ask them the meanings of the words that they did not circle.

For the translation test, it is claimed that the first language translation test should not be used in vocabulary testing because it reduces students' opportunities to practice the English language. However, the construct of the translation test in this study is the ability to translate the meanings of the target words. The translation test is a direct way of testing this construct. It allows the students to respond to vocabulary items in a way that does not draw on English language knowledge which is not directly relevant to the construct (Nation, 2001). This implies that the translation test is likely to have construct validity.

For the Laufer's (1989) formula, according to Chapter 3, the calculation was expressed in two main stages. It starts with the calculation of the percentage of unknown words that are not reported as unknown. When multiplied by the reported number of unknown words, we got the number of words that was not reported as unknown but should have been. Next, we add the number of reported words to those that should have been reported in order to calculate how many words in the text which were really unknown. Then to see how many words were known, the number of actual unknown words is subtracted from the number of words in the text. The second stage converts the number of known words into percentage. This formula tends to calculate number of actual unknown words, number of known words and the percentage of known words in a text very well in order to get TSVK. However, the formula was found to be problematic in the present study. The formula is intended to subtract the percentage of known words in a text when there is a discrepancy or when a participant did not report some words as unknown but he/she could not translate the words. However, it did not really do the job. For example, Participant 1 reported 8 unknown words in a text with 151 tokens and there was 1 discrepancy while Participants 7, 16, 26, 35 and 92 reported only one unknown word but there were eight discrepancies. Before using the formula, we might have thought that these participants might have had similar TSVK or the latter group of participants might get lower TSVK than Participant 1. Surprisingly, the participant 1, who is aware of their vocabulary knowledge, has TSVK of 94.57% while participants 7, 16, 26, 35 and 92, who reports only one unknown but could not translate 8 words in the translation test, has TSVK of 99.21%. This tends to destroy the construct validity of TSVK.

6.2 Implications for Teachers' and Students' Text Selection

As mentioned in Chapter 1, the vocabulary-based readability index was devised to measure the level or degree of the ease or difficulty of texts for Thai university students to comprehend the texts in order to help teachers and students select suitable texts for students and for themselves. Hence, some implications for text selection should be drawn in terms of teachers' and students' selection as follows.

For teachers, the investigation of the effect of LFP, vocabulary size and TSVK on reading comprehension might help raise teachers' awareness that vocabulary closely relates to reading comprehension. Furthermore, the index might be used to help teachers select texts when a placement test like DIALANG (Alderson and Huhta, 2005) is conducted before the course starts. However, the teachers are normally supposed to select texts before the course starts so the three components of the index seem to have a massive practicality problem because at the moment, they do not serve its purpose. That is to say, the readability index does not allow teachers to select texts for the students in the contexts in which they normally select texts which are before they meet the students because they do not know the students and they cannot measure their students' vocabulary size or TSVK in advance. In addition, texts used for reading instruction should not have been mastered by the students before class (Lazar, 2004). This makes TSVK inappropriate for teachers' text selection.

For students' perspectives, the index may be useful when the students would like to select an electronic text for self-study. However, it is very impractical in its current form because they are supposed to analyze LFP, complete the yes/no test or the self-report on unknown words and the translation test and use one or more of the three components of the vocabulary-based readability index by themselves. The problem is that it seems to be difficult for them to use RANGE, to look for a test, to score the tests, etc. This makes it impossible for the students to use the index. Ideally, if the index is going to be used for students' self-study, there need to be a website for dummies or a really simple straightforward website created for using it. There is an available website, namely, VocabProfile (www.lextutor.ca), which shows how a website could be designed. However, due to the constraints of the index (see 6.4), it is questionable whether the findings were strong enough to make it worth designing such a website.

Although the research findings presented in Chapter 5 showed that the indication of text difficulty by LFP seems to be as valid as the traditional readability formulas and the prediction of how easy or difficult a text is for individual students to comprehend the text seems to be promising, the index seems to be, unfortunately, far or less practical than other readability formulas such as Flesch readability formula and Flesch-Kincaid readability test. While the three components of the index requires a very tedious and detailed process with instructions that must be followed accurately, the traditional readability formulas need a very few clicks in order for the calculation to be operated.

6.3 Methodological Implications

This study has methodological implications for research on the development of readability formulas or indexes. Before the index was devised, a three-way ANOVA was used to investigate the effects of LFP, vocabulary size, TSVK on reading comprehension. It would be better to use path analysis in order to see clearer relationship between the variables so that the index would be able to predict text difficulty and reading comprehension more accurately. In order to use path analysis, a larger sample size needs to be used. According to Yamane's tables of sample size (1967), the sample size of 400 is recommended.

The benchmark for judging whether a student can be said to know words in that frequency band should be investigated because it is uncertain that the score on the yes/no test at 0.7 which was set as a benchmark for judging whether the participants know enough vocabulary on a particular frequency band to say that they know the words in the frequency band could be used. The prediction of how easy or difficult a text is for individual students in Table 5.16 revealed that texts are predicted to be optimal for the students whose vocabulary size is 0.41-0.6 or 401-600. This means that the students do not need to get 0.7 or know 700 words from a frequency band. Getting only 0.41 to 0.60 or knowing only 401 to 600 words in a frequency band is enough for them to comprehend the text containing words from that frequency band.

The yes/no test format requires the participants to circle the words that they know well enough to say what they mean. However, when a word is not circled, it is difficult to say whether it is the word that the participant read and decided not to circle or it is the word that the participant did not reach due to lack of time. It is common if the students decided not to circle the words that they did not know but the lack of time to circle all of the words they know might make score interpretation problematic. It might be better if the students were asked to write yes or y when they can recognize and know the meaning of the target word and write no or n when the word is unknown.

6.4 Constraints of the Vocabulary-based Readability Index

The vocabulary-based readability index was presumably devised in order to get rid of the drawbacks stated in 1.2. Unfortunately, these drawbacks still exist in the current form of the vocabulary-based readability index and the index also causes other drawbacks as shown below.

- According to 1.2.1, it was claimed that the traditional readability formulas tried to objectively measures readability but they ignore readers who are one of the most important factors in reading so the vocabulary-based readability index should take both text-based and reader-based variables in the development of the index. Unfortunately, the inclusion of reader-based variables especially TSVK tended to weaken the objectivity of the index because when we would like to get some information about students' TSVK, the students are supposed to circle unknown words in the self-report on unknown words. Although it is stated in the instructions that the students are required to circle the words that they do not know the meaning, it is still questionable whether the students would circle a word when they could recognize its written form but did not know its meanings. Moreover, although the translation test was used to investigate discrepancies, there were only eight words in the translation test. There is no guarantee that these words can actually detect the discrepancies.

- For the applicability of the results, the predicted reading comprehension scores were arranged into five bands of frequency: too difficult, difficult, optimal, easy and too easy in order to make the results more applicable. Unfortunately, there is a practical problem because before the bands of text difficulty can be predicted, the teachers and students are supposed to do many things.

- For the subject specific factors, the vocabulary-based readability index attempted to use TSVK as a component of the index in order to account subject specific factors. It seems to be satisfactory for most technical terms like the word "dermatitis" in 1.2.3. However, when the students find some words that have different meanings, for example, the word "accommodation" which is a technical term in law, they may not circle the word in order to report it as unknown because they may not consider the surrounding contexts and think that it means "a place to live, work, stay, etc. in" (Cambridge: Dictionaries Online, www, 2007). Hence, the use of TSVK to solve the drawback with the subject specific factors is still doubtful.

- According to 1.2, the three components of the index were claimed to objectively measure all of the four variables. Nevertheless, the calculation of some variables especially TSVK and vocabulary size in the study did not seem to be consistent. The formula for calculating TSVK is a bit complicated. If it is not calculated rigorously, the results may be varied. For the scoring of the yes/no test, as mentioned in Chapter 3, there are more than one scoring methods so when different methods are used, different scores may be obtained.

- The presentation of results of LFP seems to be less practical than the presentation of the traditional readability formulas. As mentioned in 5.2.1, the results of LFP are presented in terms percentage of tokens from the first three frequency bands. When there are two or more texts, it is easy to look at the percentage and judge which text is easier or more difficult than others. However, it seems to be impractical if a teacher or a student has only one text. For example, if he/she analyzes LFP and learns that 93% of tokens are from the first three frequency bands. It is unlikely that the percentage will make any sense for teachers or students in terms of text difficulty. For the prediction of reading comprehension scores, when a teacher or student selects a text, he/she does not think of any performance or reading scores at all so they seem to be meaningless and useless for text selection. Hence, how easy or difficulty a text is for individual students was predicted in 5.2.2 and 5.2.3 in terms of five bands of text difficulty. This kind of predictions tends to yield more practical results than the first two. However, the ranges of vocabulary size for each of the five bands of text difficulty were quite wide. A text is predicted to be optimal for the students whose vocabulary size is 401 and it is also optimal for those whose vocabulary sizes is 600. This leads to the question whether the students whose vocabulary size is 401 can comprehend a text as much as the students whose vocabulary size is 600. On the other hand, the Flesch-Kincaid readability test predicting text difficulty in terms of American grade levels was criticized in 1.2.2 that the results are not applicable to other contexts outside America. However, the results are something comprehensible that gives us some ideas whether a text should be selected. For example, a teacher of a fundamental English class consisting of non-English major students will not select a text predicted to be used for Grade 11 or 12. For a student in that class, they realize that their English language proficiency is not very high so they will not select a text for Grade 12 which is the highest predicted grade level.

- According to Chapter 3, 18 out of 120 participants were discarded from Phase 3 aiming to devise the vocabulary-based readability index because they did not report any unknown words in the self-report but they could not translate all of the target words in the translation test. This means that some participants may not be aware of their vocabulary knowledge. Fortunately, the process of getting the TSVK helped us detect such participants. However, there might have been other participants who underestimated or overestimated their own vocabulary knowledge that were not discarded from the study. The underestimation and overestimation of vocabulary knowledge might distort the TSVK and the distortion of TSVK might affect the prediction of reading comprehension at different TSVK.

- Practicality in the design of the reading comprehension test used in the development of the vocabulary-based readability index is taking priority over what should be measuring in a reading comprehension test. As mentioned in Chapter 1, Chapter 2 and Chapter 3, the reading comprehension was tested in terms of scanning and paraphrasing only. Unfortunately, scanning and paraphrasing may not be the things that make texts easy or difficult to read. The results of the index will tell anyone who uses the index whether the text that they are going to read is an easy or difficult text for them to scan or paraphrase. Therefore, the whole results are rather dubious.

- Some reading passages may contain idioms or “a group of words in a fixed order that have a particular meaning that is different from the meanings of each word understood on its own" (Cambridge Dictionaries Online, www, 2009). Some students may know the meanings of all of the individual words in an idiom but they may not know the meaning of the idiom. For instance, "touch and go", which describes a situation which is uncertain, is an example of idiom in Grant and Nation (2006) that occurs 53 times in the 100,000,000 word British National Corpus. When we look at each of the words in the idiom: "touch", "and", and "go", all of them are from the first 1000 most frequent words. When they occur very frequently, learners tend to be familiar with the words. However, the familiarity and the knowledge of these words may not help them understand the idiom because the meanings of the individual words do not relate to the idiom (Grant and Nation, 2006). Therefore, the three components of the index may not be applicable to texts with idioms.

- Since the development of the vocabulary-based readability index in Chapter 5 purely relies on the relationship between the participants' vocabulary knowledge including vocabulary size and TSVK and reading comprehension, the index tends to be applicable to only students whose vocabulary knowledge relates to their reading comprehension. It may not be applicable to some top-down and bottom-up readers who are not flexible enough to alternate between top-down and bottom-up processes according to their reading needs. On the one hand, some top-down readers may have poor vocabulary knowledge but they tend to be able to comprehend a text as a whole because they can use top-down strategies such as guessing meanings of unknown words from contexts, constructing meanings on the basis of their background knowledge and making use of text cues (Hacquebord and Stellingwerf, 2007) very well. On the other hand, some bottom-up readers mainly hold on bottom-up processing. As a result, they may concentrate on each word in a text and neglect text comprehension which is the purpose of reading comprehension because they do not seem to grasp the meaning of the text as a whole although they have good vocabulary knowledge. This may hinder them from comprehending the text. Therefore, the pure use of vocabulary knowledge to predict reading comprehension may not reflect actual reading comprehension of these top-down and bottom-up readers.

- As mentioned in Chapter 3, the analysis of LFP of reading passages and the selection of words in the yes/no test relied on word lists of BNC from spoken section only because they were the only word lists of BNC available when the LFP was analyzed and when the yes/no test was constructed (P. Scholfield, personal communication, October 15, 2006). When the latest word lists from the whole 100,000,000 British Nation Corpus was revised and sent to the researcher via e-mail by Paul Nation on June 27, 2008. These word lists were used to analyze LFP of the reading passages in the reading comprehension test. The LFP of the four reading passages based on the whole corpus is presented below.

Table 6.1: Lexical Frequency Profile of Reading Passages from Four Frequency Bands based on the Whole BNC

	Passage One	**Passage Two**	**Passage Three**	**Passage Four**
Tokens/% in the first BNC word list	138/**95.17**	127/**83.55**	109/**83.85**	110/**78.01**
Tokens/% in the second BNC word list	6/4.14	18/**11.84**	9/**6.92**	11/**7.80**
Tokens/% in the third BNC word list	1/0.69	1/0.66	8/**6.15**	17/**12.06**
Tokens/% in the fourth BNC word list	0/0.00	0/0.00	1/0.77	1/0.71
Off-list	0/0.00	6/3.95	3/2.31	2/1.42

The analyzes of LFP in Table 6.1 did not show much difference of the LFP of the first three reading passages from the LFP presented in Table 3.3. That is to say, the analysis of LFP based on the whole BNC reveals that Passage One contains 95% of tokens from the first 1000 most frequent word families, Passage Two consists of more than 95% of tokens from the first and the second 1000 most frequent word families and Passage Three contains more than 95% of tokens from the first, the second and the third 1000 most frequent word families. For Passage Four, more than 95% of tokens in the passage are from the first three frequency bands only. As a result, it might be questionable whether the results of the study and the prediction of the reading scores for Passage Four or LFP4 could be used with reading passages with all of the four word lists (the first, the second, the third and the fourth 1000 most frequent word families) or only the first, the second and the third 1000 most frequent word families. Moreover, when we look through the summaries of results for simple regression analysis for the prediction of reading comprehension scores on reading passages with LFP1, LFP2, LFP3 and LFP4 in Table 5.11, Table 5.12, Table 5.13 and Table 5.14 respectively, we can see that the a value, which is the predicted reading comprehension score when the vocabulary size equals 0, for LFP4 in Table 5.14 was the highest. Since Passage Four was supposed to be the most difficult one, the minimum predicted score should have been lower than others. This might lead to the conclusion that the predicted reading scores as well as the prediction of reading comprehension for LFP4 might not be a suitable predictor for reading passages with 95% of tokens from the first, the second, the third and the fourth 1000 most frequent word families. However, there might be some other reasons why the minimum predicted score for Passage Four is higher than others. This might be because the students had some vocabulary knowledge from the first three word lists and such knowledge helped them comprehend Passage Four although they did not know any of the word from the fourth 1000 most frequent word families in the passage.

6.5 Suggestions for Further Studies

This section presents suggestions for further studies. It aims to strengthen future research in the area of readability. The suggestions are presented as follows:

- Research will be necessary to investigate the generalisability across different samples and across different response formats or test types.

- When LFP is analyzed in order to indicate text difficulty, interpretations and descriptions of the results or the percentage of tokens from the first three frequency band should be studied in depth so as to be able to provide clear interpretations for users of the index.

- Future studies might usefully include additional variables that may affect reading comprehension such as general comprehension ability measured by listening tasks, working memory span, word recognition speed (Shiotsu, 2003 quoted in Shiotsu and Weir, 2007), syntax, schemata, phonological awareness.

- The extent that the reading models including top-down, bottom-up and interactive might affect readability and reading comprehension should be investigated.

- The inclusion of test items with other levels of cognitive domain and levels of comprehension in future studies following the research procedures of this study might also evidence an applicability and generalisability of the vocabulary-based readability index.

6.6 Summary

This study aims to devise and validate the three components of the vocabulary-based readability index. These components were: 1) the indication of text difficulty by LFP; 2) the prediction of reading comprehension at different vocabulary size and 3) the prediction of reading comprehension at different TSVK. This study is an attempt to measure readability of texts in order to get rid of the drawbacks stated in Chapter 1. After devising and validating the index in Chapter 5, the discussions of the validation of the vocabulary-based readability index in terms of face, concurrent, predictive and construct validity have been presented in this chapter. Furthermore, implications for teachers' and students' text selection and methodological implications were also presented. The chapter ends with suggestions for further studies.

REFERENCES

Aaron, P. G., and Joshi, R.M. (1992). **Reading problems: Consultation and remediation**. New York: The Guilford Press.

Abels, M. (1994). **Ken ik dit woord?** Unpublished MA Thesis, University of Nijimegen

Abram, M. J. (1981). Readability: It's use in adult education. **Lifelong Learning** 4 (5): 8-9, 30-31.

Abu-Rabia, S. (1998). Reading Arabic texts: Effects of text type, reader type, and vowelization. **Reading and Writing: An Interdisciplinary Journal** 10: 106-119.

Adams, G., and Peck, T. (2000). **101 helpful hints for IELTS: Academic Module**. Sydney: Adams & Austen.

Adams, G., and Peck, T. (2001). **202 useful exercises for IELTS: International edition**. Sydney: Adams & Austen.

Adams, M. J. (2004). Modelling connections between word recognition and reading. In R. Ruddell and N. Urnau (Eds.), **Theoretical models and processes of reading**. Newark: International Reading Association.

Aebersold, J. A. and Field, M. L. (1997). From reader to reading teacher: Issues and strategies for second language classroom. Cambridge: Cambridge University Press.

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

Alderson, J. C., and Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the common European framework. **Language Testing** 22: 301-320.

Alderson, J. C., and Urquhart, A. H. (1984). **Reading in a foreign language**. London: Longman.

Alderson, J. C., Clapham, C. and Wall, D. (1995). **Language test construction and evaluation**. Cambridge: Cambridge University Press.

Allan, S., McGhee, M., and Krieken, R. V. (2005). **Using readability formulae for examination questions** [Online]. Available: http://www.qca.org.uk/library Assets/media/allan_et_al_using_readability_formulae_for_examination_questi ons_pdf_05_1607.pdf

Al-Nujaidi, A. H. (2003). **The relationship between vocabulary size, reading strategies and reading comprehension of EFL learners in Saudi Arabia** [Online]. Available: http://e-archive.library.okstate.edu/dissertations/AAI30 94023/

Anderson, R. C., and Davison, A. (1988). Conceptual and empirical bases of readability formulas. In A. Davison and G. Green (Eds.), **Critical approaches to readability** (pp. 23-54).

Anderson, R. C., and Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (ed.). **Comprehension and teaching: Research reviews**. Newark: International Reading Association.

Anderson, R. C., and Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huxton (Ed.), **Advances in reading/language research**. Volume 2. Greenwich: JAI Press.

Astika, G. G. (1993). Analytical assessments of foreign students' writing. **RELC Journal** 24.1: 61-70.

Ayatollahi, M. A. (2003). **Effect of text structure on reading comprehension as measured by structured and unstructured summary writing**. MA thesis, Shiraz University.

Bachman, L. F., and Palmer, A. S. (1996). **Language testing in practice: designing and developing useful language tests**. Oxford: Oxford University Press.

Bailey, K. M. (1998). **Learning about language assessment: dilemmas, decisions, and directions**. London: Heinle & Helnle.

Barrow, J., Nakanishi, Y., and Ishino, H. (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. **System** 27: 223-247.

Bauer, L., and Nation, I.S.P. (1993). Word families. **International Journal of Lexicography** 6(4): 253-279.

Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., and Van de Valde, H. (2001). Examining the yes/no vocabulary test some methodological issues in theory and practice. **Language Testing** 18(3): 235-274.

Bell, T. (2006). **Active reading 1**. Singapore: Learners Publishing.

Bensoussan, M. (1984). A comparison of cloze and multiple-choice reading comprehension tests of English as a foreign language. **Langauge Testing** 1: 101-104.

Bertram, R., Baayen, R., and Schreuder, R. (2000). Effects of family size for complex words. **Journal of Memory and Language** 42: 390-405.

Bertram, R., Laine, M., and Virkkala, M. (2000). The role of derivational morphology in vocabulary acquisition: Get by with a little help from my morpheme friends. **Scandinavian Journal of Psychology** 41(4): 287-296.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (1956). **Taxonomy of educational objectives: The classification of educational goals, handbook I: Cognitive Domain**. New York: David McKay.

Bolanle, A. (2007). Good English for what?: Learners' motivation as a factor in declining learners' performance in English language acquisition and use in Nigerian schools. **Changing English** 14(2): 231-245.

Boning, R. A. (1995). **Multiple reading skills**. London: McGraw-Hill.

Botel, M. (1962). **Botel predicting readability levels**. Chicago: Follett.

Boyle, R., and Kirk, T. (2006). **Investigating the relationship between students' vocabulary knowledge and their academic reading ability**. The Seventh Annual U.A.E. University Research conference.

Brantmeier, C. (2005). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. **The Modern Language Journal** 89(i): 37-53.

Brantmeier, C. (2006). Toward a multicomponent model of interest and L2 reading: Sources of interest, perceived situational interest, and comprehension. **Reading in a Foreign Language** 18(2): 89-115.

Brisbois, J. I. (1995). Connections between first-and second-Language reading. **Journal of Reading Behavior** 27(4): 565-584.

Brown, J. D. (1996). **Testing in language programs**. Upper Saddle River, NJ: Prentice Hall Regents.

Brown, J. D. (2000). What is construct validity? **Shiken: JALT Testing & Evaluation SIG Newsletter** 4(2): 8-12.

Brown, J. D. (2005). **Testing in language programs: A comprehensive guide to English language assessment**. New York: McGraw-Hill.

Brown, K. and Hood, S. (2004). **Academic encounters: Life in society student's book.** Cambridge: Cambridge University Press.

Buck, G. (2001). **Assessing listening**. Cambridge: Cambridge University Press.

Bussayasiri, S. (1989). **English Note**. Bangkok: Modern Academic Center (MAC).

Cambridge Dictionaries Online. (2007). **Accommodation** [Online]. Available: http://dictionary.cambridge.org/dictionary/british/accommodation

Cambridge Dictionaries Online. (2009). **Idiom** [Online]. Available: http://dictionary.cambridge.org/define.asp?key=38931&dict=CALD

Cambridge Dictionaries Online. (2010). **Comprehend** [Online]. Available: http://dictionary.cambridge.org/dictionary/british/comprehend

Cambridge Dictionaries Online. (2010). **Dermatitis** [Online]. Available: http:// dictionary.cambridge.org/dictionary/british/dermatitis

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. **Language Teaching Research** 6(2): 145-173.

Carrell, P. (1984). The effects of rhetorical organization on ESL readers. **TESOL Quarterly**, 18(3): 441-469.

Carrell, P. (1985). Facilitating ESL reading by teaching text structure. **TESOL Quarterly** 19(4): 727-753.

Carrell, P. L. (1987). Readability in ESL. **Reading in a Foreign Language** 4(1): 21-40.

Carroll, J. B. (1972). Defining language comprehension: Some speculations. In R. O. Freedle and J. B. Carroll (Eds.) **Language comprehension and the acquisition of knowledge**. New York: Wiley.

Castillo, E. S. (1990). Validation of the RELC test of proficiency in English for academic purposes. **RELC Journal** 21(2): 70-86.

Cathy Duffy Reviews (2009). **McCall-Crabbs Standard Test Lessons in Reading, Books A-F** [Online]. Available: http://cathyduffyreviews.com/phonics_reading/mccall-crabbs.htm

Center for Advancement and Learning, Muskingum College. (n.d.). **Reading comprehension** [Online]. Available: http://muskingum.edu/~cal/database/general/reading.html

Chall, J. S. (1958). **Readability: An appraisal of research and application**. Columbus: Ohio State University Press.

Chall, J. S. (1984). Readability and prose composition: continuities and discontinuities. In J. Flood (ed.). **Understanding reading comprehension**. Newark: International Reading Association.

Chall, J. S. and Dale, E. (1950). Familiarity of selected health terms. **Educational Research Bulletin** 39: 197-206

Chapanis, A. (1965). Words, words, words. **Human Factors** 7(1): 1-17.

Chapanis, A. (1988). “Words, words, words” revisited. **International Review of Ergonomics** 2: 1-30.

Chujo, K., and Utiyama, M. (2005). Understanding the role of text length, sample size and vocabulary size in determining text coverage. **Reading in a Foreign Language** 17(1): 1-22.

Chung, T., and Nation, I.S.P. (2003). Technical vocabulary in specialised texts. **Reading in a Foreign Language** 15(2): 103-116.

Cobb, T. (2006). **Compleat lexical tutor** [Online]. Available: http://www.lextutor.ca/

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. **Language Learning and Technology** 11(3): 38-63.

Cook, V. (1991). **Second language learning and language teaching**. London: Edward Arnold.

Cooper, L. (2008). **Improving the vocabulary of low SES students with developmental delays: a response to intervention evaluation**. MA Thesis, Florida State University.

Cowie, A. P. (1989). **Oxford advanced learner's dictionary of current English**. Oxford: Oxford University Press.

Coxhead, A. (2000). A new academic word list. **TESOL Quarterly** 34(2): 213-238.

Dale, E., and Chall, J. (1948). A formula for predicting readability. **Educational Research Bulletin** 27: 37-54.

Daneman, M. (1991). Individual differences in reading skills. In R. Barr, M. L. Kamil, P. Mosenthal and P. D. Pearson (eds.). **Handbook of reading research: Volume II**. New York: Longman.

Davey, B. (1988). Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers. **Journal of Experimental Education** 56(2): 67-76.

Davies, F. (1995). **Introducing reading**. London: Penguin English.

Davis, F. B. (1968). Research in comprehension in reading. **Reading Research Quarterly** 4: 499-545.

Davis, F. B. (1972). Psychometric research on comprehension in reading. **Reading Research Quarterly** 7: 628-678.

Davison, A., and Kantor, R. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. **Reading Research Quarterly** 17(2): 187-209.

Day, R. and Bamford, J. (1998). **Extensive reading in the second language classroom**. Cambridge: Cambridge University Press.

Day, R., and Park, J. (2005). Developing reading comprehension questions. **Reading in a Foreign Language** 17(1): 60-73.

Degand, L., and Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. **Reading and Writing: An Interdisciplinary Journal** 15: 739-757.

Dennis, M. (1982). Imaging while reading text: A study of individual differences. **Memory and Cognition** 10(6): 540-545.

Department of Rhetoric & Writing, College of Liberal Arts, the University of Texas at Austin. (2006). **Text selection criteria** [Online]. Available: http://www.drw.utexas.edu/first-year/forum/text-selection

Douglas, D. (2000). **Assessing language for specific purposes**. Cambridge: Cambridge University Press.

DuBravac, S., and Dalle, M. (2002). Reader question formation as a tool for measuring comprehension: narrative and expository textual inferences in a Second Language. **Journal of Research in Reading** 25(2): 217-231.

Educational Testing Service. (2007a). **Tactics for TOEIC: Listening and reading tests practice test 1**. Boston: McGraw Hill.

Educational Testing Service. (2007b). **Tactics for TOEIC: Listening and reading tests practice test 2.** Boston: McGraw Hill.

Educational Testing Service. (2007c). **The official guide to the new TOEFL iBT**. Boston: McGraw Hill.

Ekwall, E. E., and Henry, I. B. (1968). How to find books children can read. **The Reading Teacher** 22: 230-232.

Elley, W. B. (1969). The assessment of readability by noun frequency counts. **Reading Research Quarterly** 4: 411-427.

Englert, C. S., and Hiebert, E. H. (1984). Children's developing awareness of text structure in expository materials. **Journal of Educational Psychology** 76: 65-74.

Eskey, D. E. (2005). Reading in a second language. In Hinkel, E. (Ed.), **Handbook of research in second language teaching and learning**. Mahwah: Lawrene Erlbaum.

Eyckmans, J. (2004). **Measuring receptive vocabulary size** [Online]. Available: http://dare.ubn.kun.nl/bitstream/2066/19469/1/19469_measrevos.pdf

Fang, Z. (2008). Going beyond the fab five: Helping students cope with the unique linguistic challenges of expository reading in intermediate grades. **Journal of Adolescent & Adult Literacy** 51(6): 476-487.

Farley, M. J., and Elmore, P.B. (1992). The relationship of reading comprehension to critical thinking skills, cognitive ability, and vocabulary for a sample of underachieving college freshmen. **Educational and Psychological Measurement** 52(4): 921-931.

Farr, J. N., Jenkins, J., and Paterson, D.G. (1951). Simplification of Flesch reading ease formula. **Journal of Applied Psychology** 35: 333-337.

Flesch, R. F. (1943). **Marks of readable style: A study in adult education**. New York: Bureau of Publications, Teachers College, Columbia University.

Flesch, R. F. (1948). A new readability yardstick. **Journal of Applied Psychology** 32(3): 221-233.

Flesch, R. F. (1979). **How to write plain English**. New York: Harper and Brothers.

Fraenkel, J. R., and Wallen, N. E. (1993). **How to design and evaluate research in education**. New York: McGraw Hill.

Fry, E. B. (1965). **Teaching faster reading: A manual**. Cambridge: Cambridge University.

Gardner, R. C., and Lambert, E. (1972). **Attitudes and motivation in second language learning**. Rowley: Newbury House.

Garrett, H. E. (1966). **Statistics in psychology and education**. New York: Longman.

Geiger, J. F. and Millis, K. K. (2004). Assessing the impact of reading goals and text structures on comprehension. **Reading Psychology** 29: 93-110.

Gerrits, M. A. (2009). **Dutch vocabulary of adult L2 learners**. MA thesis, Universiteit Van Amsterdam (University of Amsterdam).

Gilliland, J. (1972). **Readability**. London: University of London Press.

Golkar, M., and Yamini, M. (2007). Vocabulary, proficiency and reading comprehension. **The Reading Matrix** 7(3): 88-112.

Goodfellow, R., Jones, G., and Lamy M-N. (2002) **Assessing learners' writing using Lexical Frequency. ReCALL 14(1)** [Online]. Available: http:// iet.open.ac.uk/pp/r.goodfellow/ReCall01(2).htm

Grant, L., and Nation, I.S.P. (2006). How many idioms are there in English?. **ITL International Journal of Applied Linguistics** 151: 1-14.

Gregory, E. (n.d.). **Readability** [Online]. Available: http://agcomwww.tamu.edu/market/training/power/readabil.html

Gunning, R. (1952). **The technique of clear writing**. New York: McGraw-Hill.

Hacquebord, H., and Stellingwerf, B. (2007). Assessing vocabulary for the purpose of reading diagnosis. In H. Daller, J. Milton and J. Treffers-Daller, **Modelling and assessing vocabulary knowledge**. Cambridge: Cambridge University Press.

Harrison, C. (1977). Investigating the readability of texts. **Research Intelligence** 3(1): 44-47.

Harrison, C. (1980). **Readability in the classroom**. Cambridge: Cambridge University Press.

Hartley, J. (1996). Text Design. In D. Jonassen (Ed.). **Handbook of research for educational communications and technology**. pp. 795-820.

Heaton, J. B. (1975). **Writing English language tests: A practical guide for teachers of English as a second or foreign language**, London: Longman.

Herber, H. (1978). **Teaching reading in content areas**. Englewood Cliffs: Prentice-Hall.

Hiebert, E. H., Englert, C. S. and Brennan, S. (1983). Awareness of text structure in recognition and production of expository discourse. **Journal of Reading Behavior** 15(4): 63-79.

Higgins, E., and Tatham, L. (2003). **Exploring the potential of multiple-choice questions in assessment** [Online]. Available: www.celt.mmu.ac.uk/ltia/ issue4/higginstatham.shtml

Higgins, L. G. (2009). **An evaluation of the relationship between criterion-referenced competency test reading comprehension and Lexile scores and Fountas and Pinnell's guided reading levels in a Georgia Public School District**. Unpublised EdD thesis, Liberty University.

Hirsh, D., and Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts fro pleasure?. **Reading in a Foreign Language** 8: 689-696.

Holtzman, M. (2008). Demystifying application-based multiple-choice questions. **College Teaching** 56(2): 114-120.

Horst, M., Cobb, T., and Meara, P. (1998). Beyond a clockwork orange: acquiring second language vocabulary through reading. **Reading in a Foreign Language** 11(2).

How To Do Things. (2009). **How to understand reading comprehension tests** [Online]. Available: http://www.howtodothings.com/education/reading-comprehension-tests-for-children.

Hu Hsueh-chao, M., and Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. **Reading in a Foreign Language** 13(1): 403-430.

Hughes, A. (1989). **Testing for Language Teachers**. Cambridge: Cambridge University Press.

Hughes, A. (2003). **Testing for Language Teachers**. 2nd edition. Cambridge: Cambridge University Press.

Huhmann, B. A., Mothersbaugh, D.L., and Franke, G.R. (2002). Rhetorical figures in headings and their effect on text processing: the moderating role of information relevance and text length. **IEEE Transactions on Professional Communication** 45(3): 157-169.

Huibregtse, I., Admiraal, W., and Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. **Language Testing** 19(3): 227-245.

Hutchinson, T. (1997). **Lifelines: student's book intermediate level**. Oxford: Oxford University Press.

Hutchinson, T. (1998). **New hotline: intermediate level**. Oxford: Oxford University Press.

Improve-Reading-Skill.com. (2005-2007). **Reading Decoding Skills and the Elementary Reader** [Online]. Available: http://www.improve-reading-skills.com/reading_decoding_skills.htm

International Reading Association. (1994). **Resolution on the selection of reading materials** [Online]. Available: http://www.reading.org/downloads/resolutions/resolution97_selection_reading_materials.pdf

Internet Center for Management and Business Administration. (1999-2010). **Central Tendency** [Online]. Available: http://www.quickmba.com/stats/centralten/

Irene-Anna, D., Polyxeni, S., Christina, K., and Panayiota, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. **Reading Psychology** 26(1): 55-80.

Iwai, Y. (2010). Re-envisioning reading comprehension for English language learners. **The Internet TESL Journal** 16(4).

Jakeman, V., and McDowell, C. (1996). **Cambridge practice tests for IELTS 1**. Cambridge: Cambridge University Press.

Jakeman, V., and McDowell, C. (2003). **Insight into IELTS extra with answers**. Cambridge: Cambridge University Press.

Johnson, K. (2001). **An introduction to foreign language learning and teaching**. London: Longman.

Johnson, K. (n.d.). **Readability** [Online]. Available: http://www.timetabler.com/reading.html#top

Joshi, R. M., and Aaron, P. G. (2000). The component model of reading: simple view of reading made a little more complex. **Reading Psychology** 21: 85-97.

Kennedy, G. (2003). Amplifier collocations in the British National Corpus: implications for English language teaching. **TESOL Quarterly** 37(3): 467-487.

Klare, G. R. (1963). **The measurement of readability**. Ames: Iowa State University Press.

Klare, G. R. (1974-1975). Assessing readability. **Reading Research Quarterly** 10(1): 62-102.

Klare, G. R. (1976). A second look at the validity of readability formulas. **Journal of Reading Behaviour** 8: 129-152.

Kobayashi, K. (2007). The influence of critical reading orientation on external strategy use during expository text reading. **Educational Psychology** 27(3): 363-375.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: text organization and response format. **Language Testing** 19(2): 193-220.

Kobayashi, M. (2004). An investigation of method effects on reading comprehension test performance. In **the Interface between Interlanguage, Pragmatics and Assessment: Proceedings of the 3rd Annual JALT Pan-SIG Conference**. Tokyo, Japan: Tokyo Keizai University.

Laubach, R. S., and Koschnick, K. (1977). **Using readability**. New York: Reader's Press.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension?. In C. Lauren and M. Nordman (eds.). **From humans thinking to thinking machines**. Clevedon: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension?. In P. J. L. Arnaud and H. Bejoint (eds.). **Vocabulary and applied linguistics**. London: Macmillan.

Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? **RELC Journal** 25.2: 21-33.

Laufer, B. (1997). The lexical plight in second language reading: words you don’t know, words you think you know and words you can’t guess. In J. Coady and T. Huckin (Eds.). **Second language vocabulary acquisition: a rationale for pedagogy**. Cambridge University Press. pp. 20-34.

Laufer, B. and Sim, D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. **Foreign Language Annals** 18(5): 405-411.

Laufer, B., and Nation, I. S. P. (1995). Vocabulary size and use: lexical richness in L2 written production. **Applied Linguistics** 16: 307-322.

Lazar, A. M. (2004). **Learning to be literacy teachers in urban schools: stories of growth and change**. Newark: International Reading Association.

Learning Point Associates. (2004). A closer look at the five essential components of effective reading instruction: A review of scientifically based reading research for teachers [Online]. Available: http://www.learningpt.org/pdfs/literacy/components.pdf

Leech, G., Rayson, P., and Wilson, A. (2001). **Word frequencies in written and spoken English: based on the British National Corpus**. London: Longman.

Lehto, J. E., Scheinin, P., Kupiainen, S., and Hautamaki, J. (2001). National survey of reading comprehension in Finland. **Journal of Research in Reading** 24(1): 99-110.

Lenz, K. (2005). **An introduction to reading comprehension** [Online]. Available: http://www.specialconnections.ku.edu/cgi-bin/cgiwrap/specconn/main.php?cat=instruction&section=rc/main.

Levine, A., Ferenz, O., and Reves, T. (2000). EFL academic reading and modern technology: How can we turn our students into independent critical readers? **Teaching English as a Second or Foreign Language Journal** 4(4) [Online}. Available: http://www-writing.berkeley.edu/TESL-EJ/ej16/a1.html

Linderholm, T., and Van Den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository eext. **Journal of Educational Psychology** 94(4): 778-784.

Liu, N., and Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. **RELC Journal** 16(1): 33-42.

Lorch Jr., R. F., and van den Broek, P. (1997). Understanding reading comprehension: Current and future contributions of cognitive science. **Contemporary Educational Psychology** 22: 213-246.

Lorge, I. (1939). Predicting reading difficulty of selections for children. **The Elementary English Review** 16: 229-233.

Lorge, I. (1948). The Lorge and Flesch readability formulae: A correction. **School and Society** 67: 141-142.

Lunzer, E., and Garner, K. (1979). **The effective use of reading**. London: Heinemann Educational Books for the Schools Council.

Macmillan Dictionary. (2009). **Interest** [Online]. Available: http://www.macmillandictionary.com/dictionary/british/interest_44

Macmillan Dictionary. (2009). **Syntax** [Online]. Available: http://www.macmillandictionary.com/dictionary/british/syntax

Madsen, H. S. (1983). **Techniques in testing**. Oxford: Oxford University Press.

Mandernach, B. J. (2003). **Effective multiple-choice items** [Online]. Available: www.park.edu/cetl/quicktips/multiple.html

Mandler, J. (1978). A code in the node: The use of a story schema in retrieval. **Discourse Processes** 1: 14-35.

Maryland Institute College of Art. (2010). **Criteria for Text Selection** [Online]. Available: http://www.mica.edu/Research_at_MICA/Research_in_Community_Arts/Community_Arts_Convening_and_Research_Project/Criteria_for_Text_Selection.html

McCall, W. A., and Crabbs, L. M. (1925). **Standard test lessons in reading**. New York: Teachers' College, Columbia University.

McLaughlin, G. H. (1969). SMOG grading – a new readability formula**. Journal of Reading** 12: 639-646.

McLeod, J. J. (1962). The estimation of readability of nooks of low difficulty. **The British Journal of Educational Psychology** 32: 112-118.

McNamara, T. F. (2000). **Language testing**. Oxford: Oxford University Press.

Meara, P. (1992) **Yes-no English vocabulary test** [Online]. Available: http://www.lextutor.ca/tests/yes_no_eng/test_1/

Meara, P. (1993). Tintin and the world service: A look at lexical environments. In **IATEFL: Annual Conference Report**, 32-37.

Meara, P. (1996). **The vocabulary knowledge framework** [Online]. Available: http:// www.lognostics.co.uk/vlibrary/meara1996c.pdf

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. **Applied Linguistics** 26.1: 32-47.

Meara, P., and Buxton, B. (1987). An alternative to multiple choice vocabulary tests. **Language Testing** 4(2): 142-151.

Meara, P., and Jones, G. (1987). Tests of vocabulary size in English as a foreign language. **Polyglot** 8(1): 1-40.

Meara, P., and Jones. G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.). **Applied linguistics in society** (pp. 80-87). London: Centre for Information on Language Teaching and Research.

MEC Scanning Office. (1997). **MEC item analysis** [Online]. Available: http://www.utexas.edu/academic/mec/scan/scanitem.html

Mehrpour, S., and Riazi, A. (2004). The impact of text length on EFL students' reading comprehension. **Asian EFL Journal** 6(3).

Melendy, G. A. (2008). Motivating writers: The power of choice. **Asian EFL Journal** 10(3): 187-198.

Meyer, B. J. F. (2003). Text coherence and readability. **Topics in Language Disorders** 23(3): 204-224.

Mikulecky, B. S., and Jeffries, L. (1986). Reading Power. Reading: Addison-Wesley Publishing Company.

Mochida, A., and Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. **Language Testing** 23: 73-98.

Mohamad, A. (1999). What do we test when we test reading comprehension? **The Internet TESL Journal** 5(12).

Mori, S. (2002). Redefining motivation to read in a foreign language. **Reading in a Foreign Language** 14(2): 91-110.

Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. **System** 30(2): 225-35.

Nagy, W. E. (1988). **Teaching vocabulary to improve reading comprehension**. Urbana: National Council of Teachers of English.

Nagy, W. E. and Anderson, R. C. (1984). How many words are there in printed school English? **Reading Research Quarterly** 19(3): 304-330.

Nagy, W. E. and Herman, P. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. KcKeown and M. E. Curtis (Eds.). **The nature of vocabulary acquisition**. Hillsdale: Lawrence Erlbaum Associates.

Nagy, W. E., Anderson, R. Schommer, M., Scott, J.A., and Stallman, A. (1989). Morphological families in the internal lexicon. **Reading Research Quarterly** 24(3): 263-282.

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

Nation, I. S. P. (2005). **RANGE** [Computer software]. Available: http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx

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

Nation, I. S. P. and Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt and M. McCarthy (Eds.). **Vocabulary: Description, acquisition and pedagogy**. New York: Cambridge University Press.

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

Nation, I. S. P., and Coady, J. (1988) Vocabulary and reading. In R. Carter and M. McCarthy, **Vocabulary and language teaching**: 97-110.

Nation, I. S. P., and Heatley, A. (2002). **Range: A program for the analysis of vocabulary in texts** [Computer Software]. Available: http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx

National Institute of Child Health and Human Development. (2000). **Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups** (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Nurweni, A., and Read, J. (1999). The English vocabulary knowledge of Indonesian university students. **English for Specific Purposes** 18(2): 161-175.

Nuttall, C. (1982). **Teaching reading skills in a foreign language**. London: Heinemann.

Nuttall, C. (2000). **Teaching reading skills in a foreign language**. Oxford: Macmillan Heinemann ELT.

Nuttall, C. (2005). **Teaching reading skills in a foreign language**. 2nd edition, Oxford: Macmillan.

O'Malley, J. M., and Chamot, A. U. (1990). **Learning Strategies in Second Language Acquisition**. Cambridge: Cambridge University Press.

Office of Educational Assessment, University of Washington. (2005). **Item analysis** [Online]. Available: http://www.washington.edu/oea/pdfs/resources/item_ analysis.pdf

Ogle, D. M. (1986). K-W-L: A teaching model that develops active reading of expository text. **The Reading Teacher** 39(6): 564-570.

Okumura, N. (1998). The Variances in L2 reading ability in accordance with text type: Expository and narrative materials. **SURCLE** 1: 16-25.

Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham and D. Corson. (Eds.). **Encyclopedia of language and education, Volume 7: Language testing and assessment**. Dordrecht: Kluwer Academic Publishers.

Ownby, R. L. (2005). Influence of vocabulary and sentence complexity and passive voice on the readability of consumer-oriented mental health information on the internet. **AMIA 2005 Symposium Proceedings**.

Ozgungor, S., and Guthrie, J.T. (2004). Interactions among elaborative interrogation, knowledge, and interest in the process of constructing knowledge from text. **Journal of Educational Psychology** 96(3): 437-443.

Pakhare, J. (2010). **Effective teaching: Reading comprehension strategies** [Online]. Available: http://www.buzzle.com/articles/effective-teaching-readingcomprehension-strategies.html

Paribakht, T.S., and Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (Eds.), **Second language vocabulary acquisition: A rationale for pedagogy**. New York: Cambridge University Press.

Parvaresh, V., and Nemati, M. (2008). Metadiscourse and reading comprehension: The effects of language and proficiency. **Electronic Journal of Foreign Language Teaching** 5(2): 220-239.

Perkins, K. (1984). An analysis of four common item types used in testing EFL reading comprehension. **RELC Journal** 15: 29-43.

Pikulski, J. J. (2002). **Readability** [Online]. Available: http://www.eduplace.com/state/author/pikulski.pdf

Pinnell, G. S., Lyons, C. A., Deford, D. E., Bryk, A. S., and Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. **Reading Research Quarterly** 29(1): 9-38.

Pinyoanauntapong, B. (1984). **Criterion-referenced testing**. Bangkok: Odien Store.

Piromsombut, C. (2002). **Simple Items Analysis** [Computer program]. Available: http://www.thaiware.com/main/info.php?id=2704

Pollick, M. (2010). **Puppy love** [Online]. Available: http://www.wisegeek.com/what-is-puppy-love.htm

Prachyawisan, P. (2005). **Test on TOEIC grammar and reading**. Bangkok: Duangkamon Samai.

Pyrczak, F. (1972). Objective evaluation of the quality of multiple-choice test items designed to measure comprehension of reading passages. **Reading Research Quarterly** 8 (1): 62-71.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. **The Canadian Modern Language Review** 56(2): 282-307.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. **Language Learning** 52(3): 513-536.

Razi, S. (2005). A fresh look at the evaluation of ordering tasks in reading comprehension: Weighted marking protocol. **The Reading Matrix** 5(1): 1-15.

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

ReadingIsGood, Ltd. (2008). **Comprehension: A definition** [Online]. Available: http://www.readingisgood.com/2008/05/comprehension-a-definition/

RFP Evaluation Centers (2008). **Readability scores** [Online]. Available: http:// rfptemplates.technologyevaluation.com/Readability-Scores/Flesch-Reading-Ease-Readability-Score.html

Richards, J. C. and Sandy, C. (2008). Passages: Student Book 1. Cambridge: Cambridge University Press.

Richards, J. C., and Renandya, W. A. (2002). **Methodology in language teaching: an anthology of current practice**. Cambridge: Cambridge University Press.

Richards, J. C., Hull, J., and Proctor, S. (1998). **New interchange students book 3: English for international communication**. Cambridge: Cambridge University Press.

Richards, J., and Schmidt, R. (2002). **Longman dictionary of language teaching and applied linguistics**. Harlow: Longman.

Rogers, B. (2005). **World class readings 2: A reading skills text (student book)**. Seoul: McGraw Hill.

Rogers, J. R. (1962). A Formula for Predicting the Comprehension Level of Material to be Presented Orally. **The Journal of Educational Research** 56: 218-220.

Roller, C. M. (1990). The interaction between knowledge and structure variables in the processing of expository prose. **Reading Research Quarterly** 25(2): 79-89.

Rosa, M. H. (1994). Relationships between cognitive styles and reading comprehension of expository text of African American males. **Journal of Negro Education** 63(4): 546-555.

Roscoe, J. T. (1975). **Fundamental research statistics for the behavioral sciences**. New York: Rinehart and Winston.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Domic (Ed.). **Attention and Performance VI**. Hillsdale: Lawrence Erlbaum.

Sahanaya, W., Lindeck, J., and Stewart, R. (1998). **IELTS preparation and practice: reading and writing**. Melbourne: Oxford University Press.

Salager-Meyer, F. (1991). Reading expository prose at the post-secondary level: the influence of textual variables on L2 reading comprehension (a genre-based approach). **Reading in a Foreign Language** 8(1): 645-662.

Schmidt, K. M. and Embretson, S. E. (2003). Item response theory and measuring abilities. In I. B. Weiner, J. A. Schinka and W. F. Velicer (Eds.). **Handbook of psychology volume 2: Research methods in psychology**. New York: Wiley & Sons, Inc.

Schwarts, D., Sparkman, J. P., and Deese, J. (1970). The process of understanding and judgments of comprehensibility. **Journal of Verbal Learning and Verbal Behavior** 9: 87-93.

Scovell, D., Pastellas, V., and Knobel, M. (2004). **404 essential tests for IELTS academic module**. Sydney: Adams and Austen Press.

Seaman, P. (2003). **Multiple choice testing: Will it work for me?** [Online]. Available: www.lib.unb.ca/Texts/Teaching/JAN03/seaman.htm

Seliger, H. W., and Shahomy, E. (1989). **Second language research methods**. Oxford: Oxford University Press.

Sethaputra, S. (2003). **English-Thai dictionary**. Bangkok: Thai Wattanapanit.

Shahomy, E. (1984). Does the testing method make a difference? The case of reading comprehension. **Language Testing** 1: 147-170.

Sharpe, P. J. (2006). **Barron's TOEFL iBT 2006-2007**. New York: Barron's Educational Series.

Shillaw, J. (1996). **The application of Rasch modeling to yes/no vocabulary tests**. PhD dissertation, University of Wales Swansea, United Kingdom.

Shin, S. (2002). Effects of subskills and text types on Korean EFL reading scores. **Second Language Studies** 20(2): 107-130.

Shiotsu, T. (2003). **Linguistics knowledge and processing efficiency as predictors of L2 reading ability: A component skills analysis**. PhD dissertation, The University of Reading.

Shiotsu, T., and Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of second language reading comprehension test performance. **Language Testing** 24(1): 1-30.

Simon, S., Howe, L., and Kirschenbaum, H. (1972). **Values clarification: A handbook of practical strategies for teachers and students**. New York: Hart Publishing.

Sinclair, J. M. (1995). **Collins COBUILD dictionary**. London: Harper Collins.

Spache, G. (1953). A new readability formula for primary grade reading materials. **Elementary School Journal** 53: 410-413.

Srivarakan, R., and Tananuraksakul, N. (2002). **A needs survey of English teachers and students to further their graduate studies in the field of teaching English**. South-East Asia University, Bangkok.

Stahl, S.A. (1986). Three principles of effective vocabulary instruction. **Journal of Reading** 29(7): 662-668.

Stavans, A. and Oded, B. (1993). Assessing EFL reading comprehension: The case of Ethiopian learners. **System** 21(4): 481-494.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M.G. McKeown and M.E. Curtis (Eds.) **The nature of vocabulary acquisition**. Hillsdale: Erlbaum.

Taberski, S. (2000). **On solid ground: strategies for teaching reading K-3**. Portsmouth: Heinemann.

Taylor, M. C., and Wahlstrom, M. W. (1999). **Readability as applied to an ABE assessment instrument** [Online]. Available: http://www.nald.ca/library/research/report4/rep31-35/rep34-01.htm

Taylor, W. (1953). Cloze procedure: A new tool for measuring readability. **Journalism Quarterly** 30: 415-433.

Thiengburanathum, W. (1998). **SE-Ed's modern English-Thai fictionary**. Bangkok: SE-Education.

Thorndike, E. L., and Lorge, I. (1944). **The teacher's word book of 30,000 words**. New York: Bureau of Publications, Teachers College, Columbia University.

Tinker, M. A. (1963). **Legibility of print**. Ames: Iowa State University Press.

Trosborg, A. (1997). Text typology: register, genre and text type. **Text Typology and Translation**: 3-23.

U.S. Department of Education. (2004). **Three questions on the importance of reading** [Online]. Available: http://www.ed.gov/news/newsletters/extracredit/2004/0202.html

University of Cambridge Local Examination Syndicate. (2000). **Cambridge IELTS 2**. Cambridge: Cambridge University Press.

University of Cambridge Local Examination Syndicate. (2002). **Cambridge IELTS 3**. Cambridge: Cambridge University Press.

University of Oxford. (2005). **British National Corpus** [Online]. Available: http://www.natcorp.ox.ac.uk/corpus/index.xml

University of Prince Edward Island. (2006). **Glossary of terms** [Online]. Available: http://www.upei.ca/~xliu/measurement/glossary.htm

Urbaniak, G. C., and Plous, S. (2007). **Research randomizer** [Online]. Available: http://www.randomizer.org/form.htm

Valcourt, G., and Wells, L. (1999). **Mastery: A university word list reader** MI: University of Michigan Press.

Varnhagen, C. K. (1991). Text relations and recall for expository prose. **Discourse Processes** 14: 399-422.

Vaso, G. (2005). **Determining the readability of a book** [Online]. Available: http://www.eclectichomeschool.org/articles/article.asp?articleid=429&resourceid=2

Victorian Association for the Teaching of English. (2008). **VATE text selection guidelines** [Online]. Available: http://www.vate.org.au/vate.nsf/Weben/Text%20Selection%20Guidelines

Wallace, C. (1992). **Reading**. Oxford: Oxford University Press.

Wan-a-rom, U. (2008). Comparing the vocabulary of different graded-reading schemes. **Reading in a Foreign Language** 20(1): 43-69.

Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. **Journal of Verbal Learning and Verbal Behavior** 9: 398-404.

Watts, L., and Nisbet, J. (1974). **Legibility in children’s book – A review of research**. Windsor: NFER.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. **Studies in Second Language Acquisition** 30(1): 79-95.

Wen, X. (1997). Motivation and language learning with students of Chinese. **Foreign Language Annals** 30: 235-250.

Werlich, E. (1976). **A text grammar of English**. Heidelberg: Quelle & Meyer.

West, R. F., and Stanovich, K.E. (1991). The incidental acquisition of information from reading. **Psychological Science** 2(5): 325-329.

Wiersma, W. and Jurs, S. (2005). **Educational measurement and testing**. Englewood Cliffs: Prentice Hall.

Wikipedia (2010). **Reading comprehension** [Online]. Available: http://en.wikipedia.org/wiki/Reading_comprehension

Wikipedia (2010). **Spearman rho** [Online]. Available: http://en.wikipedia.org/wiki/Spearman_Rho

Wilhelm, J. (1996-2010). **Understanding reading comprehension** [Online]. Available: http://www2.scholastic.com/browse/article.jsp?id=4465

Williams, D. (1984). **Reading in the language classroom**. London: Macmillian.

Yamane, T. (1967). **Statistics: An introductory analysis**. New York: Harper and Row.

Yap, K. O. (1979). Vocabulary – Building blocks of comprehension?. **Journal of Reading Behavior** 1: 49-59.

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

APPENDIX A

Yes/No Test

วงกลมรอบคำศัพท์ที่นักศึกษาสามารถบอกความหมายได้ (Circle the words that you know well enough to say what they mean.)

Level 1

abogative	trouble	troake	finance	nickling
pring	great	treat	hubbard	health
court	contrivial	shoe	general	believe
overend	society	quiet	suggest	lader
stace	dogmatile	cantileen	motion	responsible
garden	join	require	local	reduce
opie	complete	haque	lannery	loveridge
lower	gume	community	baldock	couple
buttle	final	wray	space	improve
accept	cleaner	kitchen	danger	galpin

Level 2

fear	violent	pocock	ordinary	quorant
attach	charlet	accuracy	connect	tooley
benevolate	dirty	hoult	revolution	fortnight
spread	whaley	smile	kiley	mundy
balfour	mental	wake	swim	invite
satisfy	suffer	chop	ralling	vene
chief	degate	replace	length	desire
tough	twose	hapgood	rudge	moffat
bodelate	avoid	direction	culture	attitude
indicate	burn	duffin	limidate	pray

Level 3

external	sympathy	oven	smooth	acklon
whitrow	gentle	detailoring	lamb	fatal
pauling	vickery	flame	escrotal	collapse
accuse	devil	horozone	breath	bance
reward	recenticle	horror	channing	fiction
webbert	nonagrate	crazy	dowrick	courage
whistle	litter	justal	soul	adair
gallery	scudamore	raw	flood	suddery
enemy	extract	merry	contortal	manner
outcome	interfere	litholect	personality	batcock

Level 4

preserve	jealous	submarine	cheerful	berrow
classify	explore	connery	cambule	horobin
eluctant	reform	perfume	bastionate	esteem
heritage	wildle	gorgeous	pursue	jungle
ripe	crocodile	draconite	reservory	ambitious
privacy	scurrilize	modify	gender	charactal
cottonwool	jarvis	aistrope	fluctual	staple
refund	accompany	glandle	shallow	profile
condimented	almanical	outstanding	herb	absent
grove	pernicate	conclude	urban	eckett

APPENDIX B

Translation Test for the Pilot Work of the Yes/No Test

แปลคำศัพท์ด้านล่างเป็นภาษาไทยหรือภาษาอังกฤษ (Translate the following words into Thai or English)

1. finance ____
2. final ____
3. treat ____
4. local ____
5. suggest ____
6. general ____
7. reduce ____
8. responsible ____
9. invite ____
10. accuracy ____
11. fortnight ____
12. satisfy ____
13. burn ____
14. desire ____

15. indicate ______________________________

16. tough ______________________________

17. sympathy ______________________________

18. manner ______________________________

19. outcome ______________________________

20. interfere ______________________________

21. accuse ______________________________

22. horror ______________________________

23. raw ______________________________

24. courage ______________________________

25. classify ______________________________

26. heritage ______________________________

27. absent ______________________________

28. cheerful ______________________________

29. gender ______________________________

30. herb ______________________________

31. shallow ______________________________

32. ambitious ______________________________

APPENDIX C

Vocabulary Knowledge Scale

วงกลมเลือกหมายเลข 1-4 ซึ่งบ่งบอกถึงความรู้ที่มีต่อคำศัพท์แต่ละคำ

quiet

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

garden

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

local

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

health

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

burn

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

invite

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

ordinary

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

drag

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

soul

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

extract

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

manner

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

merry

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

jungle

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

grove

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

elegant

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

spontaneous

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

brilliant

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

cheap

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

couple

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

community

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

nasty

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

grace

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

tough

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

holy

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

donate

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

elbow

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

greedy

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

trophy

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

ripe

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

heave

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

legislate

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

disruption

1. ฉันไม่เคยเห็นคำนี้มาก่อนเลย
2. ฉันเคยเห็นคำนี้ แต่ไม่รู้ว่ามันแปลว่าอะไร
3. ฉันเคยเห็นคำนี้ แต่คิดว่ามันแปลว่า .... (คำแปลหรือคำเหมือน)
4. ฉันรู้จักคำนี้ มันหมายความว่า .... (คำแปลหรือคำเหมือน)

Vocabulary Knowledge Scale

Choose the appropriate number from 1 to 4 that represent your knowledge of each word.

quiet

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

garden

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

local

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

health

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

burn

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

invite

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

ordinary

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

drag

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

soul

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

extract

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

manner

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

merry

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

jungle

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

grove

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

elegant

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

spontaneous

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

brilliant

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

cheap

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

couple

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

community

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

nasty

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

grace

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

tough

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

holy

1. I don’t remember having seen this word before.

2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

donate

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

elbow

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

greedy

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

trophy

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before, and I think it means _________ (synonym or translation).
4. I know this word. It means _________ (synonym or translation).

ripe

1. I don’t remember having seen this word before.
2. I have seen this word before, but I don’t know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

heave

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

legislate

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

disruption

1. I don't remember having seen this word before.

2. I have seen this word before, but I don't know what it means.

3. I have seen this word before, and I think it means _________ (synonym or translation).

4. I know this word. It means _________ (synonym or translation).

APPENDIX D

The Reading Comprehension Test with Four Passages

PASSAGE ONE

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

ข้อที่ 1-6

The sixth sense is usually called ‘common sense’. It is not like the other senses because it is not linked to a part of the body, such as seeing with the eyes, smelling with the nose and hearing with the ears. It could have something to do with the brain, but it is not something we can learn at school.

Common sense is what we use when we decide what to do in a certain situation, for example, what do we do when we lock ourselves out of the house? or how do we drive when the road is wet and dangerous?

Common sense does not usually come from the heart. So it is not very useful when one has to decide about a personal matter, for example, in choosing a marriage partner. However, not everyone seems to have the same amount of common sense.

Questions 1-6

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

1. Why is common sense not like the other senses?
    a. Because it has nothing to do with a part of the body.
    b. Because we can learn it at school.
    c. Because different people have different ideas about it.
    d. Because it is linked to a part of the body.

2. Common sense is probably linked to _____________.
    a. the eyes
    b. the brain
    c. the heart
    d. the ears

3. How many parts of the body linked to the other senses are mentioned in the passage?
    a. One
    b. Two
    c. Three
    d. Four
4. Which of the following is not mentioned in the passage?
    a. Common sense
    b. Sixth sense
    c. Extrasensory perception
    d. None of the above
5. According to the passage, which of the following statements is true?
    a. We learn common sense at school.
    b. Everyone uses common sense to choose a marriage partner.
    c. Everyone uses common sense to make decisions.
    d. We have different amounts of common sense.
6. Which of these requires common sense?
    a. Solving a mathematical problem
    b. Reading the newspaper
    c. Solving an everyday problem
    d. Choosing a husband or wife

PASSAGE TWO

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

ข้อที่ 7-12

Can you get a cold if the weather changes suddenly, when you are exposed to a draught or if you go outdoors too soon after washing your hair? According to the Contact Coldtime Survey, that's what most people believe. The truth, however, is that none of these factors influence your risk of getting sick. The most common way that can put you in jeopardy, according to the U.S. Department of Health and Human Services, is touching an object after someone with a cold had touched it. Other ways include; being too near when someone coughs or sneezes, close indoor proximity to someone who has a cold, and not getting enough sleep.

Which sex complains more when sick? Only 43% of males admitted to the survey they complain and whine more than females. However, many women will tell you (and 70% did) that it's a man who complains most with a cold.

Questions 7-12

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

7. According to the passage, what was the purpose of the survey?
    a. To find out how people catch colds.
    b. To establish how people believe they can catch cold.
    c. To learn whether men catch more colds than women.
    d. To find out how people caught colds last year.
8. Who conducted the survey stating how people believe they can catch cold?
    a. A women's magazine
    b. The Contact Coldtime Company
    c. A men's magazine
    d. The U.S. Department of Health and Human Services

9. According to the passage, why are you more likely to get a cold?
    a. You go out when your hair is wet.
    b. You are not used to the sudden changes of the weather.
    c. You are exposed to a draught.
    d. You do not get enough sleep.
10. What should you do if you do not want to get a cold?
    a. Stay at home when the weather changes.
    b. Do not go out after washing your hair.
    c. Stay away from someone who coughs or sneezes.
    d. Do not go out when it is windy.

11. Which statement is most accurate according to the passage?
    a. Children get more colds.
    b. Men admitted they complain more when they get a cold.
    c. Women admitted they complain more when they get a cold.
    d. Men and woman say children complain more when they get a cold.
12. According to the passage, which statement is **<u>NOT</u>** the factor influencing your risk of getting sick?
    a. Sitting very close to your friend with a cold.
    b. Going out when your hair is wet.
    c. Being too near to your friend who is coughing or sneezing.
    d. Touching something after your friend who caught a cold had touched it.

PASSAGE THREE

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

ข้อที่ 13-18

The kindergarten child will develop many interests in the world around him and the teacher can use field trips to the school grounds or around the school to generate curiosity to learn new things. For example, a story about pets, plus a walk around the school with the teacher pointing out the things described in a story, can generate ideas for many pictures done with crayons or paints. The child is thus developing ways of communicating. He uses his picture and talks about his experience.

The regular routine in kindergarten helps children establish patterns of living. Going to the toilet, washing hands, and getting drinks help in establishing habits of orderly living. These regular habits are part of maturity development which prepares a child for his next stage of growth.

Questions 13-18

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

13. The main idea of the first paragraph concerns ______________.
    a. the development of kindergarten children
    b. communication of kindergarten children
    c. the maturation of children
    d. living patterns of children
14. The field trips referred to in the passage would be ______________.
    a. time consuming
    b. in the school grounds
    c. in the next stage of growth
    d. far from the school

15. Habits established in kindergarten should ______________.
    a. be changed by the parents
    b. be carefully monitored by the teacher
    c. generate ideas about life
    d. get a child ready for the next growth period
16. According to the passage, the field trips ______________.
    a. arouse the child's interest to learn new things around them.
    b. encourage the child to be independent.
    c. promote the child's social skills.
    d. establish the child's patterns of living.
17. According to the passage, which statement is true?
    a. The child is healthy as he walks around the schoolyard.
    b. The field trip is a block to progress in child development.
    c. The child's regular routine helps form habits of living.
    d. The field trip should be done when the child is mature.
18. How many regular habits are mentioned in the passage?
    a. One
    b. Two
    c. Three
    d. Four

PASSAGE FOUR

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

ข้อที่ 19-24

The government of Australia is pleased with new statistics revealing a decline in the nation's trade deficit.

The Australian Trade Committee (ATC) reveals that the nation's exports totalled $72.8 billion in May which is an increase of 8 percent over last year. Exports to the European Union increased by 0.7 percent. A 10 percent rise in exports of heavy machinery goods, such as semiconductors and steel products, to developing nations was an important note in the study.

Imports were only 2 percent higher, reaching $194.3 billion. Imports from the European Union and Japan steadily declined over the past year. However, the opposite is true of trade with the United States with which imports have increased by 6 percent. Analysts believe that the main reason for the decrease in the trade deficit has been a decline in purchases and prices for petroleum products from the Middle East.

Questions 19-24

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

19. What is the purpose of the study?

    a. To please the government of Australia.
    b. To study the exports and imports of the country.
    c. To export heavy machinery goods.
    d. To find out the reason for the decrease of purchases of petroleum products.

20. What did the study find?
    a. The country's trade balance has improved.
    b. Australia has had a trade surplus for three consecutive years.
    c. The country's trade agreement with Japan has helped the textile industry.
    d. Trade barriers with European nations must be removed to reduce the trade deficit.
21. In what industry was the largest growth in exports shown?
    a. Textiles
    b. Agriculture
    c. Heavy machinery
    d. Mineral exploration
22. According to the passage, which of the following statements is true?
    a. Exports from the European Union have reached $72.8 billion.
    b. New statistics on the Australian trade deficit were revealed by the Australian Trade Committee.
    c. Most of the semiconductors were exported to the European Union and Japan.
    d. The government of Australia banned exports of heavy machinery goods to developing nations.
23. According to the passage, what happened to imports from the United States?
    a. Imports of steel products from the United States have increased.
    b. There was about 6 percent increase in imports from the United States.
    c. The European Union and Japan opposed imports from the United States.
    d. Imports from the United States have reached $194.3 billion.
24. According to the passage, which statement is **<u>NOT</u>** the reason for the decrease of the trade deficit?
    a. The nation's exports have been increased.
    b. The economic crisis in Australia has got worse.
    c. Semiconductors and steel products have been sold to developing countries.
    d. Purchases of petroleum products have declined.

APPENDIX F

The Translation Test for Testing Text-Specific Vocabuıary Knowledge in the Reading Comprehension Test with Four Passages

แปลคำที่ขีดเส้นใต้เป็นภาษาไทยหรือภาษาอังกฤษ **(Translate the underlined words into Thai or English.)**

1. What do we do when we lock **ourselves** out of the house?

2. Common sense is what we use when we decide what to do in a **certain** situation.

3. However, not everyone seems to have the same **amount** of common sense.

4. Common sense is what we use when we decide what to do in a certain **situation**.

5. Common sense is what we use when we **decide** what to do in a certain situation.

   So it is not very useful when one has to **decide** about a personal matter, for example, in choosing a marriage partner.

6. So it is not very useful when one has to decide about a personal matter, for example, in choosing a **marriage** partner.

7. It could have something to do with the **brain**, but it is not something we can learn at school.

____

8. Common sense does not usually come from the heart. So it is not very useful when one has to decide about a personal matter, for example, in **choosing** a marriage partner.

____

9. The most common way that can put you in **jeopardy**, according to the U.S. Department of Health and Human Services, is touching an object after someone with a cold had touched it.

____

10. Can you get a cold if the weather changes suddenly, when you are exposed to a **draught** or if you go outdoors to soon after washing your hair?

____

11. Other ways include; being too near when someone coughs or **sneezes**, close indoor proximity to someone who has a cold, and not getting enough sleep.

____

12. Other ways include; being too near when someone coughs or sneezes, close indoor **proximity** to someone who has a cold, and not getting enough sleep.

____

13. The truth, however, is that none of these factors **influence** your risk of getting sick.

____

14. Can you get a cold if the weather changes suddenly, when you are **exposed** to a draught or if you go outdoors to soon after washing your hair?

____

15. Which sex **complains** more when sick? While only 43% of males admitted to the survey they **complain** and whine more than females. Many women will tell you (and 70% did) that it's a man who **complains** most with a cold.

____

16. Other ways include; being too near when someone **coughs** or sneezes, close indoor proximity to someone who has a cold, and not getting enough sleep.

____

17. The regular routine in kindergarten helps children **establish** patterns of living. Going to the toilet, washing hands, and getting drinks help in **establishing** habits of orderly living.

____

18. For example, a story about buildings of pets, plus a walk around the schoolyard with the teacher pointing out the things described in a story, can generate ideas for many pictures done with **crayons** or paints.

____

19. The **kindergarten** child will develop many interests in the world around him and the teacher can use field trips to the school grounds or around the block to arouse interest to the point of having projects grow out of these experiences.

____

20. These regular habits are part of **maturity** schedule which prepares a child for his next stage of growth.

____

21. The kindergarten child will develop many interests in the world around him and the teacher can use field trips to the school grounds or around the block to **arouse** interest to the point of having projects grow out of these experiences.

____

22. The regular **routine** in kindergarten helps children establish patterns of living.

____

23. These regular habits are part of maturity **schedule** which prepares a child for his next stage of growth.

____

24. For example, a story about buildings of pets, plus a walk around the schoolyard with the teacher pointing out the things described in a story, can **generate** ideas for many pictures done with crayons or paints.

____

25. A 10 percent rise in exports of heavy machinery goods, such as **semiconductors** and steel products, to developing nations was an important note in the study.

____

26. Analysts believe that the main reason for the decrease in the trade deficit has been a decline in **purchases** and prices for petroleum products from the Middle East.

____

27. The government of Australia is pleased with new statistics revealing a decline in the nation's trade **deficit**.

Analysts believe that the main reason for the decrease in the trade **deficit** has been a decline in purchases and prices for petroleum products from the Middle East.

____

28. A 10 percent rise in exports of heavy **machinery** goods, such as semiconductors and steel products, to developing nations was an important note in the study.

____

29. The government of Australia is pleased with new statistics **revealing** a decline in the nation's trade deficit.

____

30. Analysts believe that the main reason for the decrease in the trade deficit has been a decline in purchases and prices for **petroleum** products from the Middle East.

____

31. The government of Australia is pleased with new statistics revealing a **decline** in the nation's trade deficit.

    Analysts believe that the main reason for the decrease in the trade deficit has been a **decline** in purchases and prices for petroleum products from the Middle East.

    ____

32. Imports from the European Union and Japan **steadily** declined over the past year.

    ____

APPENDIX G

The Reading Comprehension Test with Five Passages from TOEIC Tests

PASSAGE ONE

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

Questions 1-4

Town Meeting to Discuss Tunnel Project

The Clarion City Council will hold a meeting this week to discuss the proposal to build a tunnel under the Central River. The tunnel would not replace the existing bridge but would relieve traffic jams on the roads that link the city center to suburbs on the south side of the river. Yanco Builders, seen as the firm likely to be given the building contract, has been working with council members ever since preliminary studies of the tunnel project began. A spokesperson for Yanco, Donald Young, stated, “This tunnel will have a tremendous impact on the people of Clarion. Traffic conditions will improve; noise pollution will be reduced, and the struggling shops and restaurants in the city center will see a growth in business. If people in the suburbs can get to the city center more easily, they will go there for their shopping and entertainment.”

Some opponents of the tunnel project question whether the tunnel will actually reduce traffic jams. A citizens group calling itself Responsible Traffic Solutions (RTS) has been pushing for a temporary delay in the project so that further studies can be conducted. RTS, which has requested a meeting with city council members, believes that the city’s shortage of public transportation might be a more important issue to address when considering how to reduce traffic on the bridge. The group has also expressed concern over the loss of green space along the river. Says an RTS spokesperson, “To take away the little green space the city has would be harmful to plant and animal species that live along the southern edge of the city and would take away from the natural charm of the area.”

Questions 1-4

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

1. What is true about the city council?

   a. It has voted to approve the tunnel project.

   b. It has been working with Yanco Builders.

   c. It plans to build more shops and restaurants.

   d. It has met with the leaders of RTS.

2. What did Mr. Young say about the tunnel project?

   a. It will be completed later than expected.

   b. It will create more jobs in the suburbs.

   c. It will replace the existing bridge.

   d. It will be good for the economy in the city center.

3. Why has RTS asked that the project be delayed?

   a. It wants citizens to vote on the project.

   b. It wants to extend the tunnel to a neighboring town.

   c. It wants more studies to be done.

   d. It wants a different company to build tunnel.

4. What is **NOT** stated about the city of Clarion?

   a. Its shops have recently seen a growth in business.

   b. Traffic is a problem between the city center and the suburbs.

   c. Its public transportation system is insufficient.

   d. Its green space is very limited.

PASSAGE TWO

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

Questions 5-9

Classy Shopping Bags Winning Customers

Consumers' excessive use of hard-to-recycle plastic bags is harmful to the environment. Supermarkets have used many tactics to encourage their customers not to use these plastic bags. Some try charging extra for each bag used or offering paper bags instead, but these methods are far from perfect. Charging for a plastic bag can have the effect of driving customers to a competing supermarket. Paper bags are expensive to make, so they are less viable than even low-quality plastic bags. Many supermarkets offer a small discount in return for not using a plastic bag, and still others sell reusable cloth sacks that customers can bring back every time they come shopping. But the availability of cloth bags has done little to reduce the use of plastic disposables. Marketing experts at Ipanerra Supermarkets Company believe cloth bags would be more successful if they were more fashionable.

According to them, it is hard to blame customers for not buying a durable cloth carryall, even a cheap one, if it doesn't have a catchy design, and if all it features is a company logo.

To address this problem, Ipanerra hired a team of designers to make its cloth bags more colorful than those available at other supermarkets. Rather than sporting company logos, Ipanerra bags feature pictures of animals, drawings of famous buildings, and cartoon figures commissioned for this purpose. The result? Not only is there considerable demand for the bags but they also attract attention in the street, which means free advertising for the supermarket. The bags are easily associated with the supermarket chain even without the logos, which Ipanerra removed from both its cloth and plastic bags. The success has prompted the company to consider designing small sacks suited for younger customers, a logical next step for a supermarket chain targeting children.

Questions 5-9

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

5. What method of reducing the use of plastic bags is **NOT** mentioned?
   a. Introducing extra charges for plastic bags
   b. Providing customers with other kinds of bags
   c. Offering a discount for not using plastic bags
   d. Publicizing the harmful effects of using plastic bags

6. According to the article, why are paper bags **NOT** a good alternative?
   a. It is difficult to recycle them.
   b. They contain harmful substances.
   c. It costs a lot to produce them.
   d. They are generally of low quality.

7. What is the disadvantage of many supermarkets' cloth bags?
   a. Their durability b. Their appearance
   c. Their cost d. Their size

8. What did Ipanerra stop printing on its bags?
   a. Pictures of animals b. Company logos
   c. Drawings of buildings d. Cartoon figures

9. What does Ipanerra plan to do next?
   a. Offer bags for children
   b. Reduce the price of its cloth bags
   c. Advertise its bags in the media
   d. Increase the size of its cloth bags

PASSAGE THREE

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

Questions 10-12

The Starfish Water Park is one of the greatest attractions for both the tourists and residents of the sea resort town of Zopot. Before this impressive swimming pool complex was inaugurated last May, it had generated enough public interest for its owners to dispense with expensive advertising campaigns. While new facilities of this kind are often recommended by famous people on television prior to opening, all that the water park needed was word of mouth for business to take off. Residents of Zopot had known about the construction since it started in October two years ago, and had anxiously waited for the opening day.

But for all the water park's success, its owners had worried that the pool might be situated too close to the beach. And in fact, the complex received fewer visitors last August, when many tourists chose to swim in the nearby sea. But the situation changed completely in December, a record-breaking month, when outdoor temperatures dropped, and entire families waited in half-hour lines to enter the recreational complex. This is understandable given the variety and quality of the facilities, including pools with wonderful simulated waves and long, roller-coaster-like slides.

Ticket prices range from \$8 to \$20 depending on the time of day; general admission tickets allow access to all sections of the water park except the sauna area, which costs and additional \$5.

Questions 10-12

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

10. What does the passage imply about the opening day of the water park?
    - a. It did not generate much interest.
    - b. It was not advertised on television.
    - c. It had to be postponed by one month.
    - d. It was attended by famous people.

11. What were people concerned about regarding the water park?
    - a. The choice of location
    - b. The cost of construction
    - c. The building design
    - d. The safety code

12. When did the water park receive the most visitors?
    - a. In May
    - b. In August
    - c. In October
    - d. In December

PASSAGE FOUR

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

Questions 13-16

Instruction in Photography

Garnello Studio is pleased to offer instruction in photography in three different areas as outlined below. Serious amateurs are invited to join studio owner Steve Garinello for comprehensive, fun lessons in photography. Though class sizes vary, only ONE photographer at a time shoots pictures during each session. Classes are offered Monday through Friday except where noted otherwise.

Portrait Session

Participants will work with several different styles of lighting and learn how to achieve the best exposures. Instruction will be given in how best to pose the sitter, as well as how to integrate different props and backgrounds. We usually work through four different settings in a session, and sessions are limited to ten photographers. 7.30 to 10.00 p.m.; £25 per participant.

Fashion Session

In addition to the key areas of lighting, poses, props, and backgrounds, participants will be given instruction in how to direct a model and what techniques can be used to bring out the model's best shots. Again, we typically work through four different scenes, and sessions are limited to ten participants. 7.30 to 10.00 p.m.; £28 per participant.

Advertising and Still Life Session

This is an introduction to the principles of advertising design. Special attention will be paid to the positioning and presentation of the product, as well as any models required in the shot. Some shots will be based on actual advertisements created by Steve Garinello, while others will be created to illustrate a specific technique. Sessions are limited to six photographers. 7.00 to 11.00 p.m.; £30 per participant.

All-Day Session

On Saturdays we offer full-day instruction covering Portrait and Fashion photography. Sessions are limited to eight photographers. 9.30 a.m. to 3.00 p.m.; £60 (includes lunch and refreshments).

Questions 13-16

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

13. For whom is the instruction intended?

a. People who want to have their photographs taken

b. Professional photographers who are preparing exhibits

c. People who want to learn about photography

d. Models who are trying to expand their portfolios

14. How many scenes are included in a fashion session?

a. 1 b. 4

c. 6 d. 10

15. What is **NOT** mentioned as a topic in the advertising session?

a. Product presentation b. Design principles

c. Model placement d. Camera selection

16. In what session is the studio owner’s work used?

a. Portrait
b. Fashion
c. Advertising
d. All-day

PASSAGE FIVE

อ่านเนื้อเรื่องด้านล่าง และวงกลมคำศัพท์ที่นักศึกษาไม่สามารถบอกความหมายได้ และตอบคำถาม

(Read the passage below, circle the words that you do not know the meanings and answer the questions.)

Questions 17-20

Housing Business Recovered

Canada’s housing market is showing signs of gradual recovery. After struggling for nearly nine years against a protracted economic recession, housing is finally showing a substantial amount of growth. According to Canada Mortgage and Housing Corporation (CMHC), lower mortgage rates and prices will boost construction of new homes by 11 percent this year and 9 percent next year. In addition, sales of existing homes are expected to jump 14 percent over the course of the year. Despite these encouraging figures, the market is still much weaker than it was a decade ago. CMHC economist, Pat Trent, says slow population growth, high employment rates and declining real incomes are largely to blame for the situation.

Questions 17-20

เลือกคำตอบที่ถูกต้องที่สุดและเขียนคำตอบลงในกระดาษคำตอบ (Choose the best answer and write your answer on the answer sheet.)

17. What is being reported in the passage?

a. An improvement in Canada’s housing situation

b. The rate of return on real estate investments

c. A rise in the number of homeless people over the decade

d. How the economic recession affected emigration rates from Canada to the United States

18. What is expected to increase by nine percent in the next year?

a. Mortgage rates　　b. Employment rates

c. Sales of existing homes　　d. Construction of new homes

19. How has the market changed from the previous decade?

a. It has become weaker.

b. It has become stronger.

c. Nothing has changed.

d. It has experienced domestic growth but international decline.

20. According to Pat Trent, what has caused the overall market trend?

a. A reduction in trade barriers

b. Speculation by Canadian economics

c. Increased costs of building supplies

d. High employment rates and slow population growth

APPENDIX H

The Translation Test for Testing Text-Specific Vocabulary Knowledge in the Reading Comprehension Test with Five Passages from TOEIC Tests

แปลคำที่<u>ขีดเส้นใต้</u>เป็นภาษาไทยหรือภาษาอังกฤษ (Translate the underlined words into Thai or English.)

1. The Clarion City Council will hold a meeting this week to discuss the proposal to build a **<u>tunnel</u>** under the Central River. The **<u>tunnel</u>** would not replace the existing bridge but would relieve traffic jams on the roads that link the city center to suburbs on the south side of the river.

---

2. The tunnel would not replace the existing bridge but would **<u>relieve</u>** traffic jams on the roads that link the city center to suburbs on the south side of the river.

---

3. Yanco Builders, seen as the firm likely to be given the building contract, has been working with council members ever since **<u>preliminary</u>** studies of the tunnel project began.

---

4. A spokesperson for Yanco, Donald Young, stated, “This tunnel will have a **<u>tremendous</u>** impact on the people of Clarion.

---

5. Traffic conditions will improve; noise pollution will be reduced, and the **<u>struggling</u>** shops and restaurants in the city center will see a growth in business.

---

6. Some **opponents** of the tunnel project question whether the tunnel will actually reduce traffic jams.

7. A citizens group calling itself Responsible Traffic Solutions (RTS) has been pushing for a **temporary** delay in the project so that further studies can be conducted.

8. RTS, which has requested a meeting with city council members, believes that the city's **shortage** of public transportation might be a more important issue to address when considering how to reduce traffic on the bridge.

9. **Classy** Shopping Bags Winning Customers

10. Paper bags are expensive to make, so they are less **viable** than even low-quality plastic bags.

11. Many supermarkets offer a small discount in return for not using a plastic bag, and still others sell reusable cloth **sacks** that customers can bring back every time they come shopping.

12. But the availability of cloth bags has done little to reduce the use of plastic **disposables**.

13. According to them, it is hard to blame customers for not buying a **durable** cloth carryall, even a cheap one, if it doesn't have a catchy design, and if all it features is a company logo.

14. According to them, it is hard to blame customers for not buying a durable cloth **carryall**, even a cheap one, if it doesn't have a catchy design, and if all it features is a company logo.

____

15. According to them, it is hard to blame customers for not buying a durable cloth carryall, even a cheap one, if it doesn't have a **catchy** design, and if all it features is a company logo.

____

16. Rather than sporting company logos, Ipanerra bags feature pictures of animals, drawings of famous buildings, and cartoon figures **commissioned** for this purpose.

____

17. The Starfish Water Park is one of the greatest **attractions** for both the tourists and residents of the sea resort town of Zopot.

____

18. Before this impressive swimming pool complex was **inaugurated** last May, it had generated enough public interest for its owners to dispense with expensive advertising campaigns.

____

19. Before this impressive swimming pool complex was inaugurated last May, it had **generated** enough public interest for its owners to dispense with expensive advertising campaigns.

____

20. Before this impressive swimming pool complex was inaugurated last May, it had generated enough public interest for its owners to **dispense** with expensive advertising campaigns.

____

21. Residents of Zopot had known about the **construction** since it started in October two years ago, and had anxiously waited for the opening day.

____

22. And in fact, the **complex** received fewer visitors last August, when many tourists chose to swim in the nearby sea.

____

23. But the situation changed completely in December, a record-breaking month, when outdoor temperatures dropped, and entire families waited in half-hour lines to enter the **recreational** complex.

____

24. This is understandable given the variety and quality of the facilities, including pools with wonderful **simulated** waves and long, roller-coaster-like slides.

____

25. Serious **amateurs** are invited to join studio owner Steve Garinello for comprehensive, fun lessons in photography.

____

26. Serious amateurs are invited to join studio owner Steve Garinello for **comprehensive**, fun lessons in photography.

____

27. Participants will work with several different styles of lighting and learn how to achieve the best **exposures**.

____

28. Instruction will be given in how best to pose the **sitter**, as well as how to integrate different props and backgrounds.

____

29. Instruction will be given in how best to pose the sitter, as well as how to **integrate** different props and backgrounds.

___

30. Instruction will be given in how best to pose the sitter, as well as how to integrate different **props** and backgrounds.

___

31. Some shots will be based on actual advertisements created by Steve Garinello, while others will be created to **illustrate** a specific technique.

___

32. Sessions are limited to eight photographers. 9.30 a.m. to 3.00 p.m.; £60 (includes lunch and **refreshments**).

___

33. Canada's housing market is showing signs of **gradual** recovery.

___

34. Canada's housing market is showing signs of gradual **recovery**.

___

35. After struggling for nearly nine years against a **protracted** economic recession, housing is finally showing a substantial amount of growth.

___

36. After struggling for nearly nine years against a protracted economic **recession**, housing is finally showing a substantial amount of growth.

___

37. After struggling for nearly nine years against a protracted economic recession, housing is finally showing a **substantial** amount of growth.

___

38. According to Canada **<u>Mortgage</u>** and Housing Corporation (CMHC), lower **<u>mortgage</u>** rates and prices will boost construction of new homes by 11 percent this year and 9 percent next year.

____

39. According to Canada Mortgage and Housing Corporation (CMHC), lower mortgage rates and prices will **<u>boost</u>** construction of new homes by 11 percent this year and 9 percent next year.

____

40. CMHC economist, Pat Trent, says slow population growth, high employment rates and **<u>declining</u>** real incomes are largely to blame for the situation.

____

APPENDIX K

Questionnaire for Investigating Predictive Validity of the Prediction of Reading Comprehension

แบบสอบถาม

วงกลมรอบตัวเลขเพื่อเลือกระดับความยากง่ายของแต่ละเนื้อเรื่อง (หมายเลข 1 = ยากเกินไป (too difficult) 2 = ยาก (difficult) 3 = เหมาะสม/กำลังดี (optimal) 4 = ง่าย (easy) 5 = ง่ายเกินไป (too easy))

	Too Difficulty	**Difficulty**	**Optimal**	**Easy**	**Too Easy**
Passage One	1	2	3	4	5
Passage Two	1	2	3	4	5
Passage Three	1	2	3	4	5
Passage Four	1	2	3	4	5

ขอบคุณที่กรุณาตอบแบบสอบถาม

Questionnaire

Circle a number in order to indicate the level of text difficulty of each reading passage (1 = too difficult, 2 = difficult, 3 = optimal, 4 = easy, 5 = too easy).

	Too Difficulty	**Difficulty**	**Optimal**	**Easy**	**Too Easy**
Passage One	1	2	3	4	5
Passage Two	1	2	3	4	5
Passage Three	1	2	3	4	5
Passage Four	1	2	3	4	5

Thank you for your cooperation

APPENDIX I

Questionnaire for Investigating Face Validity of the Three Components of Vocabulary-based Readability Index

แบบสอบถาม

วงกลมรอบตัวเลขเพื่อแสดงความคิดเห็นเกี่ยวกับความเหมาะสมของตัวชี้วัดในดัชนีวัดความยากง่ายในการอ่าน

	เห็นด้วยอย่างยิ่ง	**เห็นด้วย**	**ไม่แน่ใจ**	**ไม่เห็นด้วย**	**ไม่เห็นด้วยอย่างยิ่ง**
สัดส่วนของคำศัพท์ที่เกิดขึ้นถี่มากหรือถี่น้อย	5	4	3	2	1
คลังคำศัพท์	5	4	3	2	1
เปอร์เซ็นต์ของความรู้คำศัพท์ในแต่ละบทอ่าน	5	4	3	2	1

ขอบคุณที่กรุณาตอบแบบสอบถาม

Questionnaire

Circle a number to indicate the extent that you agree that each indicator of text difficulty indicates text difficulty or predict reading comprehension.

	Extremely agree	**Agree**	**Unsure**	**Disagree**	**Extremely disagree**
Lexical frequency profile	5	4	3	2	1
Vocabulary size	5	4	3	2	1
Text-specific vocabulary knowledge	5	4	3	2	1

Thank you for your cooperation.

APPENDIX J

Questionnaire for Investigating Concurrent Validity of LFP as an Indicator of Text Difficulty

แบบสอบถาม

เรียงลำดับเนื้อเรื่องทั้ง 4 เรื่อง (Passage One, Passage Two, Passage Three และ Passage Four) ที่ได้อ่านมาแล้ว ตามลำดับความยากง่าย (หมายเลข 1 = ง่ายที่สุด หมายเลข 4 = ยากที่สุด)

______________ Passage One

______________ Passage Two

______________ Passage Three

______________ Passage Four

ขอบคุณที่กรุณาตอบแบบสอบถาม

Questionnaire

Number the four reading passages (Passage One, Passage Two, Passage Three, Passage Four) from 1 to 4 in terms of text difficulty (1 = the easiest 4 = the most difficult).

______________ Passage One

______________ Passage Two

______________ Passage Three

______________ Passage Four

Thank you for your cooperation

APPENDIX L

Scores on the Yes/No Test from the Pilot Work

of the Yes/No test

Participant	Score on Level 1	Score on Level 2	Score on Level 3	Score on Level 4	Total Score
A1	0.87	0.80	0.58	0.45	0.67
A2	0.82	0.40	0.38	0.05	0.42
A3	0.40	0.13	0.00	0.00	0.13
A4	1.00	0.75	0.90	0.73	0.85
A5	0.58	0.20	0.10	0.03	0.23
A6	0.63	0.07	0.03	0.00	0.18
A7	0.43	0.03	0.00	0.00	0.12
A8	0.37	0.10	0.03	0.00	0.12
A9	0.87	0.30	0.10	0.07	0.33
A10	0.77	0.13	0.10	0.07	0.27
A11	0.90	0.70	0.37	0.27	0.56
A12	0.70	0.10	0.07	0.00	0.22
A13	1.00	0.97	0.80	0.60	0.84
A14	0.80	0.33	0.07	0.03	0.31
A15	0.45	0.13	0.00	0.08	0.17
A16	0.58	0.10	0.10	0.07	0.21
A17	0.60	0.18	0.13	0.10	0.25
A18	0.65	0.10	0.07	0.03	0.21
A19	0.67	0.07	0.07	0.13	0.23
A20	0.72	0.7	0.20	0.02	0.30
A21	0.73	0.17	0.07	0.03	0.25
A22	0.92	0.77	0.63	0.30	0.65
A23	0.53	0.20	0.10	0.05	0.22
A24	0.60	0.30	0.32	0.12	0.33
A25	0.47	0.10	0.03	0.00	0.15
A26	0.73	0.27	0.13	0.07	0.30
A27	0.70	0.13	0.07	0.10	0.25
A28	0.30	0.10	0.00	0.02	0.10
A29	0.53	0.17	0.03	0.07	0.20
A30	0.67	0.13	0.02	0.00	0.20
A31	0.92	0.80	0.52	0.03	0.57
A32	0.83	0.33	0.17	0.08	0.35
A33	0.68	0.20	0.12	0.02	0.25
A34	0.60	0.10	0.07	0.03	0.20

Participant	Score on Level 1	Score on Level 2	Score on Level 3	Score on Level 4	Total Score
A35	0.80	0.45	0.53	0.35	0.53
A36	0.67	0.10	0.10	0.03	0.22
A37	0.42	0.17	0.03	0.03	0.16
A38	0.87	0.43	0.30	0.03	0.41
A39	0.85	0.50	0.43	0.50	0.57
A40	0.82	0.53	0.70	0.10	0.54
A41	0.87	0.83	0.78	0.57	0.76
A42	0.45	0.17	0.13	0.17	0.23
A43	0.42	0.12	0.10	0.20	0.21
A44	0.55	0.20	0.30	0.23	0.32
A45	0.20	0.07	0.00	0.03	0.08
A46	0.67	0.17	0.03	0.07	0.23
A47	0.45	0.10	0.03	0.08	0.17
A48	0.78	0.58	0.52	0.47	0.59
A49	0.65	0.25	0.07	0.27	0.31
A50	0.23	0.13	0.03	0.12	0.13
A51	0.27	0.10	0.03	0.08	0.12
A52	0.97	0.67	0.37	0.37	0.59
A53	0.50	0.23	0.07	0.08	0.22
A54	0.90	0.63	0.53	0.43	0.62
A55	0.77	0.23	0.03	0.07	0.27
A56	0.65	0.23	0.20	0.07	0.29
A57	0.77	0.30	0.20	0.37	0.41
A58	0.77	0.10	0.10	0.13	0.28
A59	0.77	0.23	0.33	0.30	0.41
A60	0.67	0.27	0.10	0.22	0.31
A61	0.90	0.45	0.45	0.48	0.57
A62	0.67	0.13	0.10	0.15	0.26
A63	0.77	0.38	0.18	0.25	0.40
A64	0.73	0.13	0.07	0.23	0.29
A65	0.53	0.0	0.10	0.13	0.22
A66	0.83	0.90	0.72	0.50	0.74
A67	0.60	0.27	0.13	0.27	0.32
A68	0.82	0.27	0.13	0.05	0.32
A69	0.90	0.30	0.33	0.40	0.48
A70	0.50	0.10	0.07	0.10	0.19
A71	0.73	0.27	0.33	0.13	0.37
A72	0.30	0.00	0.00	0.03	0.08
A73	0.27	0.10	0.17	0.27	0.20
A74	0.73	0.37	0.47	0.18	0.44
A75	0.73	0.13	0.22	0.23	0.33

APPENDIX M

Results from Questionnaire for Investigating Face Validity in the Pilot Work of the Reading Comprehension Test

Items	Teachers					ΣR	Mean
	A	B	C	D	E		
1	+1	+1	+1	+1	+1	5	1.0
2	+1	+1	+1	+1	+1	5	1.0
3	+1	+1	+1	+1	+1	5	1.0
4	+1	+1	+1	+1	+1	5	1.0
5	+1	+1	+1	+1	+1	5	1.0
6	+1	0	+1	+1	+1	4	0.8
7	0	0	+1	+1	+1	3	0.6
8	+1	+1	+1	+1	+1	5	1.0
9	+1	0	+1	+1	+1	4	0.8
10	+1	0	+1	+1	+1	4	0.8
11	+1	+1	+1	+1	+1	5	1.0
12	+1	+1	+1	+1	+1	5	1.0
13	0	0	+1	+1	+1	3	0.6
14	+1	0	+1	+1	+1	4	0.8
15	+1	+1	+1	+1	+1	5	1.0
16	+1	+1	+1	+1	+1	5	1.0
17	+1	+1	+1	+1	+1	5	1.0
18	+1	+1	+1	+1	+1	5	1.0
19	+1	0	0	+1	+1	3	0.6
20	+1	+1	+1	+1	+1	5	1.0
21	+1	+1	+1	+1	+1	5	1.0
22	+1	+1	+1	+1	+1	5	1.0
23	+1	+1	+1	+1	+1	5	1.0
24	+1	+1	+1	+1	+1	5	1.0

APPENDIX N

Scores on the Four Reading Passages with Four Different LFP in the Reading Comprehension Test

Participants	LFP1	LFP2	LFP3	LFP4
1	5.0	3.0	3.0	2.0
2	5.0	5.0	4.0	3.0
3	1.0	1.0	0.0	2.0
4	4.0	1.0	1.0	3.0
5	5.0	5.0	1.0	2.0
6	6.0	6.0	5.0	3.0
7	4.0	3.0	3.0	4.0
8	5.0	5.0	2.0	2.0
9	6.0	6.0	5.0	4.0
10	5.0	6.0	5.0	3.0
11	5.0	5.0	5.0	3.0
12	6.0	6.0	6.0	6.0
13	6.0	6.0	5.0	5.0
14	3.0	0.0	0.0	2.0
15	3.0	0.0	0.0	2.0
16	6.0	5.0	5.0	3.0
17	4.0	3.0	0.0	2.0
18	6.0	6.0	6.0	3.0
19	4.0	1.0	0.0	2.0
20	3.0	0.0	0.0	1.0
21	5.0	4.0	4.0	3.0
22	5.0	5.0	2.0	2.0
23	2.0	2.0	2.0	1.0
24	6.0	5.0	4.0	3.0
25	4.0	4.0	5.0	3.0
26	6.0	4.0	5.0	4.0
27	6.0	5.0	5.0	4.0
28	4.0	4.0	4.0	3.0
29	4.0	3.0	3.0	3.0
30	4.0	1.0	2.0	2.0
31	6.0	6.0	5.0	3.0
32	2.0	1.0	0.0	2.0
33	5.0	4.0	1.0	2.0
34	6.0	3.0	3.0	2.0
35	5.0	5.0	3.0	2.0

Participants	LFP1	LFP2	LFP3	LFP4
36	6.0	4.0	3.0	4.0
37	3.0	1.0	0.0	2.0
38	5.0	5.0	3.0	2.0
39	4.0	4.0	2.0	3.0
40	1.0	1.0	0.0	2.0
41	4.0	5.0	0.0	2.0
42	6.0	6.0	4.0	4.0
43	4.0	4.0	3.0	3.0
44	4.0	0.0	0.0	2.0
45	4.0	5.0	3.0	1.0
46	5.0	1.0	0.0	2.0
47	1.0	1.0	0.0	1.0
48	5.0	4.0	2.0	2.0
49	6.0	4.0	2.0	1.0
50	5.0	5.0	4.0	3.0
51	5.0	5.0	3.0	3.0
52	3.0	2.0	2.0	2.0
53	5.0	4.0	3.0	2.0
54	6.0	6.0	4.0	5.0
55	4.0	2.0	2.0	2.0
56	4.0	3.0	3.0	1.0
57	2.0	2.0	3.0	1.0
58	5.0	3.0	3.0	3.0
59	5.0	5.0	3.0	3.0
60	6.0	3.0	3.0	3.0
61	4.0	4.0	3.0	3.0
62	5.0	4.0	4.0	3.0
63	3.0	2.0	2.0	2.0
64	1.0	1.0	0.0	1.0
65	3.0	2.0	1.0	1.0
66	5.0	4.0	3.0	1.0
67	6.0	2.0	2.0	1.0
68	5.0	4.0	3.0	1.0
69	3.0	1.0	1.0	1.0
70	5.0	3.0	2.0	1.0
71	3.0	3.0	2.0	1.0
72	3.0	2.0	1.0	0.0
73	4.0	1.0	1.0	0.0
74	4.0	4.0	3.0	2.0
75	6.0	5.0	4.0	2.0
76	3.0	2.0	2.0	1.0
77	5.0	5.0	3.0	4.0
78	5.0	4.0	3.0	5.0
79	5.0	4.0	2.0	2.0
80	3.0	1.0	1.0	0.0

Participants	LFP1	LFP2	LFP3	LFP4
81	3.0	3.0	1.0	1.0
82	2.0	2.0	1.0	0.0
83	5.0	4.0	3.0	1.0
84	3.0	3.0	2.0	2.0
85	4.0	2.0	2.0	0.0
86	4.0	3.0	3.0	1.0
87	3.0	3.0	1.0	0.0
88	5.0	4.0	3.0	3.0
89	4.0	3.0	2.0	1.0
90	3.0	3.0	2.0	0.0
91	4.0	3.0	3.0	2.0
92	6.0	5.0	5.0	4.0
93	4.0	4.0	4.0	3.0
94	3.0	3.0	2.0	0.0
95	4.0	4.0	2.0	2.0
96	4.0	4.0	2.0	0.0
97	5.0	4.0	5.0	3.0
98	6.0	5.0	5.0	4.0
99	5.0	3.0	4.0	2.0
100	5.0	4.0	5.0	4.0
101	6.0	6.0	6.0	6.0
102	4.0	4.0	5.0	4.0

APPENDIX O

Vocabulary Size Scores

Level 1

Participants	Real Words	Pseudowords	Scores
1	29	4	0.77
2	22	0	0.73
3	10	1	0.28
4	27	3	0.75
5	24	2	0.70
6	30	2	0.90
7	17	1	0.52
8	26	3	0.72
9	29	2	0.87
10	24	0	0.80
11	27	1	0.85
12	27	0	0.90
13	30	1	0.95
14	15	1	0.45
15	21	1	0.65
16	26	0	0.87
17	16	0	0.53
18	29	1	0.92
19	27	4	0.70
20	14	1	0.42
21	23	0	0.77
22	21	0	0.70
23	13	1	0.38
24	28	1	0.88
25	26	1	0.82
26	29	1	0.92
27	28	1	0.88
28	24	1	0.75
29	27	3	0.75
30	18	2	0.50
31	30	1	0.95
32	16	3	0.38
33	25	0	0.83
34	29	1	0.92
35	24	1	0.75

Participants	Real Words	Pseudowords	Scores
36	30	1	0.95
37	24	3	0.65
38	29	4	0.77
39	23	2	0.67
40	8	0	0.27
41	24	1	0.75
42	27	0	0.90
43	23	0	0.77
44	22	3	0.58
45	22	0	0.73
46	24	2	0.70
47	12	2	0.30
48	21	0	0.70
49	30	0	1.00
50	24	1	0.75
51	30	3	0.85
52	15	2	0.40
53	27	1	0.85
54	29	1	0.92
55	25	0	0.83
56	20	0	0.67
57	19	2	0.53
58	26	3	0.72
59	24	1	0.75
60	27	1	0.85
61	20	0	0.67
62	25	2	0.73
63	19	0	0.63
64	14	0	0.47
65	17	2	0.47
66	23	1	0.72
67	28	1	0.88
68	25	0	0.83
69	14	0	0.47
70	25	2	0.73
71	19	1	0.58
72	22	4	0.53
73	28	1	0.88
74	27	3	0.75
75	28	1	0.88
76	18	0	0.60
77	25	0	0.83
78	25	0	0.83
79	23	2	0.67
80	22	1	0.68

Participants	Real Words	Pseudowords	Scores
81	25	3	0.68
82	13	0	0.43
83	23	1	0.72
84	22	1	0.68
85	20	2	0.57
86	18	2	0.50
87	16	2	0.43
88	27	3	0.75
89	19	2	0.53
90	13	0	0.43
91	16	0	0.53
92	30	2	0.90
93	27	1	0.85
94	14	0	0.47
95	27	1	0.85
96	24	2	0.70
97	30	3	0.85
98	29	1	0.92
99	21	0	0.70
100	28	1	0.88
101	30	0	1.00
102	27	2	0.80

Level 2

Participants	Real Words	Pseudowords	Scores
1	16	2	0.43
2	24	2	0.70
3	9	1	0.25
4	13	3	0.28
5	24	2	0.70
6	28	1	0.88
7	14	0	0.47
8	22	1	0.68
9	26	0	0.87
10	28	3	0.78
11	28	2	0.83
12	29	1	0.92
13	29	1	0.92
14	2	0	0.07
15	9	3	0.15
16	26	3	0.72
17	19	3	0.48

Participants	Real Words	Pseudowords	Scores
18	27	0	0.90
19	9	2	0.20
20	5	2	0.07
21	22	0	0.73
22	21	0	0.70
23	9	0	0.30
24	25	0	0.83
25	23	0	0.77
26	25	2	0.73
27	27	1	0.85
28	24	2	0.70
29	18	1	0.55
30	13	3	0.28
31	30	2	0.90
32	8	1	0.22
33	22	3	0.58
34	22	0	0.73
35	22	1	0.68
36	24	0	0.80
37	9	1	0.25
38	21	0	0.70
39	23	0	0.67
40	9	1	0.25
41	22	1	0.68
42	29	1	0.92
43	17	0	0.57
44	4	1	0.08
45	21	0	0.70
46	8	1	0.22
47	12	2	0.30
48	20	0	0.67
49	19	0	0.63
50	29	4	0.77
51	22	1	0.68
52	13	1	0.38
53	27	2	0.80
54	26	0	0.87
55	20	0	0.67
56	17	2	0.47
57	15	0	0.50
58	13	0	0.43
59	25	2	0.73
60	20	0	0.67
61	19	2	0.53
62	23	2	0.67

Participants	Real Words	Pseudowords	Scores
63	19	1	0.58
64	9	0	0.30
65	18	3	0.45
66	19	0	0.63
67	16	4	0.33
68	22	1	0.68
69	16	2	0.43
70	25	3	0.68
71	19	1	0.58
72	14	0	0.47
73	16	2	0.43
74	24	1	0.75
75	22	0	0.73
76	16	1	0.48
77	28	3	0.78
78	26	1	0.82
79	24	3	0.65
80	16	1	0.48
81	18	2	0.50
82	16	3	0.38
83	21	1	0.65
84	22	1	0.68
85	14	2	0.37
86	14	0	0.47
87	11	0	0.37
88	23	2	0.67
89	17	2	0.47
90	17	3	0.42
91	16	1	0.48
92	28	3	0.78
93	21	1	0.65
94	17	4	0.37
95	23	0	0.77
96	15	0	0.50
97	24	1	0.75
98	25	0	0.83
99	24	3	0.65
100	28	2	0.83
101	30	1	0.95
102	22	0	0.73

Level 3

Participants	Real Words	Pseudowords	Scores
1	12	0	0.40
2	17	0	0.57
3	9	4	0.10
4	13	3	0.28
5	8	3	0.12
6	24	1	0.75
7	21	5	0.45
8	15	2	0.40
9	28	3	0.78
10	25	2	0.73
11	21	0	0.70
12	30	1	0.95
13	28	3	0.78
14	4	2	0.03
15	5	2	0.07
16	28	3	0.78
17	5	1	0.12
18	24	1	0.75
19	7	2	0.13
20	1	0	0.03
21	20	0	0.67
22	13	2	0.33
23	2	0	0.07
24	28	3	0.78
25	21	0	0.70
26	21	0	0.70
27	23	0	0.77
28	21	1	0.65
29	18	3	0.45
30	6	3	0.05
31	28	3	0.78
32	5	1	0.12
33	12	2	0.30
34	22	1	0.68
35	27	5	0.65
36	22	0	0.73
37	8	2	0.17
38	21	1	0.65
39	27	5	0.65
40	2	0	0.07
41	8	2	0.17
42	20	0	0.67
43	15	0	0.50

Participants	Real Words	Pseudowords	Scores
44	2	0	0.07
45	21	1	0.65
46	8	3	0.12
47	9	4	0.10
48	9	0	0.30
49	10	0	0.33
50	22	0	0.73
51	24	3	0.65
52	15	3	0.35
53	23	1	0.72
54	25	2	0.73
55	21	2	0.60
56	15	1	0.45
57	11	0	0.37
58	20	5	0.42
59	22	1	0.68
60	16	1	0.48
61	20	3	0.52
62	16	0	0.53
63	16	1	0.48
64	9	2	0.20
65	13	1	0.38
66	19	2	0.53
67	10	1	0.28
68	27	5	0.65
69	13	2	0.33
70	24	3	0.65
71	15	1	0.45
72	13	0	0.43
73	15	2	0.40
74	16	0	0.53
75	23	1	0.72
76	14	0	0.47
77	25	2	0.73
78	22	0	0.73
79	18	0	0.60
80	11	2	0.27
81	15	1	0.45
82	13	1	0.38
83	14	0	0.47
84	15	0	0.50
85	13	3	0.28
86	13	1	0.38
87	12	1	0.35
88	16	1	0.48

Participants	Real Words	Pseudowords	Scores
89	12	1	0.35
90	10	1	0.28
91	14	0	0.47
92	21	1	0.65
93	19	1	0.58
94	8	0	0.27
95	16	1	0.48
96	5	1	0.12
97	21	0	0.70
98	23	0	0.77
99	19	0	0.63
100	23	0	0.77
101	27	0	0.90
102	20	0	0.67

Level 4

Participants	Real Words	Pseudowords	Scores
1	3	2	0.00
2	24	5	0.55
3	1	0	0.03
4	12	6	0.10
5	9	5	0.05
6	21	1	0.65
7	9	4	0.10
8	11	3	0.20
9	23	1	0.72
10	16	3	0.38
11	21	5	0.45
12	28	1	0.88
13	24	1	0.75
14	4	2	0.03
15	2	1	0.02
16	9	0	0.30
17	1	0	0.03
18	26	3	0.72
19	11	7	0.02
20	0	0	0.00
21	12	0	0.40
22	6	3	0.05
23	4	2	0.03
24	23	2	0.67
25	19	0	0.63

Participants	**Real Words**	**Pseudowords**	**Scores**
26	20	1	0.62
27	24	3	0.65
28	13	0	0.43
29	15	1	0.45
30	0	0	0.00
31	23	1	0.72
32	4	2	0.03
33	10	3	0.18
34	14	2	0.37
35	12	3	0.25
36	21	1	0.65
37	0	0	0.00
38	11	1	0.32
39	10	1	0.28
40	13	3	0.28
41	3	2	0.00
42	20	0	0.67
43	11	1	0.32
44	6	3	0.05
45	8	1	0.22
46	3	2	0.00
47	4	1	0.08
48	6	3	0.05
49	9	2	0.20
50	19	1	0.58
51	11	0	0.37
52	15	3	0.35
53	18	0	0.60
54	22	1	0.68
55	18	3	0.45
56	6	1	0.15
57	8	0	0.27
58	14	1	0.42
59	23	5	0.52
60	13	0	0.43
61	12	0	0.40
62	20	5	0.42
63	13	1	0.38
64	5	1	0.12
65	14	2	0.37
66	7	1	0.18
67	4	0	0.13
68	10	2	0.23
69	10	1	0.28
70	22	3	0.58

Participants	Real Words	Pseudowords	Scores
71	13	1	0.38
72	16	2	0.43
73	12	0	0.40
74	16	3	0.38
75	16	3	0.38
76	13	0	0.43
77	22	4	0.53
78	23	1	0.72
79	14	3	0.32
80	0	0	0.00
81	13	1	0.38
82	11	0	0.37
83	10	1	0.28
84	15	2	0.40
85	5	0	0.17
86	15	4	0.30
87	12	1	0.35
88	16	1	0.48
89	9	2	0.20
90	5	1	0.12
91	15	2	0.40
92	16	0	0.53
93	14	0	0.47
94	14	5	0.22
95	12	1	0.35
96	9	5	0.05
97	11	0	0.37
98	19	0	0.63
99	10	0	0.33
100	20	0	0.67
101	28	1	0.88
102	21	1	0.65

APPENDIX P

30 Texts

Text 1

The United States does not have a national university, but the idea has been around for quite some time. George Washington first recommended the idea to Congress; he even selected an actual site in Washington, D.C., and then left an endowment for the proposed national university in his will. During the century following the Revolution, the idea of a national university continued to receive the support of various U.S. presidents, and philanthropist Andrew Carnegie pursued the cause at the beginning of the present century. Although the original idea has not yet been acted upon, it continues to be proposed in bills before Congress.

Text 2

One common element in architectural design is the bracket. This simple device is found on many styles of buildings. More than merely decorative, brackets carry the weight of a canopy, porch or small balcony. Their purpose is twofold: to allow a horizontal shelter to project over a door or window and to provide more support on the wall below these structures.

The word bracket can be traced to the Latin "bracchium" meaning arm. Brackets were first introduced into architectural design in the mid 1800s by Italian builder, George Canelli who wanted to improve the strength of hotel terraces. By the turn of the century architects had expanded the practical use of these structures and were developing the ornamental features of the devices to improve the aesthetic appearance of private homes.

Text 3

It is difficult to imagine a group of people turning away from the affluence of the West and choosing to brave the harsh elements of nature and the bitterly cold, long winters of the Arctic. The Nenets, however, have done just this and have managed to preserve their culture over the past thousand years of dramatic global change.

The Nenets, direct descendants of Far East aboriginal tribes, are one of the world's last Arctic nomads. In the winter, the Nenets live in camps near the OB River in Russia and every March they begin a 900 kilometer migration across ice and tundra to the Yamal Peninsula on the northwestern tip of Siberia. They keep herds of reindeer for transportation, hides, and food.

Although they have survived the tests of the 20th century, many researchers are worried that as the nomads become increasingly subjected to western culture, modern technology and environmental degradation, the existence of this Arctic tribe will be threatened.

Text 4

Dunnlow to Deliver New Trains for Gaelic Railways

Gaelic Railways (GR), the national railway, announced yesterday that it has entered into a contract with train manufacturer Dunnlow for the purchase of fourteen new electric trains. The decision by GR's management came after a test run from Rexford to Donnebrook revealed that the new trains can cover the 45 kilometers separating the two towns in half the time it takes GR's current trains.

The new trains will gradually replace those currently in service, introduced by GR 25 years ago. The railway company had originally intended to complete the project in a 5-year span so as to let the last phase of the transition coincide with the end of the 30-year life expectancy of GR's current fleet. However, company officials were so excited by the performance of the new trains that the decision was made to pull all current trains from the tracks within 3 years. They plan to upgrade the longest routes first, beginning with the distances longer than 100 kilometers.

Negotiations between GR and Dunnlow have been going on for nearly 2 years. The lengthy duration of the talks was mostly due to reluctance on the part of the government to help fund the project. Says Shane O'Farrel, GR's representative, "We had difficulty convincing the authorities that buying new, faster trains, while expensive, was a worthwhile investment. Gradually, however, the government realized that adequate transportation is crucial for the continued development and prosperity of the region.

Text 5

New Danger to Your Heart

A recent study indicates that low level inflammation can exist for years inside arteries and can be as powerful a heart attack trigger as too much cholesterol. Researchers studied men who have experienced low level inflammation of the arteries for several years, and found them to be three times as likely to suffer heart attack and twice as likely to have strokes as normal men. Earlier studies found evidence of inflammation in heart attack sufferers, but it was not known until recently that men who appeared outwardly to be in good health could have chronic arterial inflammation.

Unfortunately, the cause of arterial inflammation remains a mystery. The most popular theory suggests it may be the result of a chronic infection caused by a common bacteria or virus that the body is unable to eradicate.

Text 6

Epinephrine or adrenaline is a natural secretion of the adrenal glands in the human body. Its primary function in the human body is to assist the body in coping with sudden surges of stress. When a person unexpectedly finds himself in a stressful situation filled with fear or anger, a large amount of epinephrine is released into the blood and the body responds with an increased heartbeat, higher blood pressure, and conversion of glycogen into glucose for energy to enable the body to deal with the stress.

It is possible to extract epinephrine from the adrenal glands of animals or to synthesize it chemically in order to put it to further use. It is used in the treatment of severe asthma, where it relaxes the large muscles of the bronchi, the large air passages leading into the lungs. It is also used in cases of severe allergic reaction or cardiac arrest.

Text 7

A massive banking crisis occurred in the United States in 1933. In the two preceding years, a large number of banks had failed, and fear of lost savings had prompted many depositors to remove their funds from banks. Problems became so serious in the state of Michigan that Governor Comstock was forced to declare a moratorium on all banking activities in the state on February 14, 1933. This panic quickly spread to other states, and on March 6, President Roosevelt declared a banking moratorium throughout the United States that left the entire country without banking services.

Congress immediately met in a special session to solve the banking crisis and on March 9 passed the Emergency Banking Act of 1933 to assist financially healthy banks to reopen. By March 15, banks controlling 90 percent of the country's financial reserves were again open for business.

Text 8

Two Big Nations To Be Linked

An ambitious project is now underway to link a fiber-optic cable between Britain and Japan. If successful, this sophisticated cable link will facilitate rapid transfer of huge amounts of data between the two nations. The cable will be installed on the ocean floor and extend a distance of 28,000 km. The cost of this mega-project will be 182.6 billion dollars but will allow 4.7 gigabytes of information to travel between the two countries per second. Sterling Communication Company, the largest investor in this project, described the many advantages that the fiber optic cable information routes have over transoceanic copper wires and satellite transmissions, including faster, more reliable and less expensive voice and data transfers. However, project coordinators hasten to add that compared to satellites, these cables will require much more regular maintenance and repair due to corrosion and abrasion.

Text 9

Manic depression is another psychiatric illness that mainly affects mood. A patient suffering from this disease will alternate between periods of manic excitement and extreme depression, with or without relatively normal periods in between. The changes in mood suffered by a manic-depressive patient go for beyond the day-to-day mood changes experienced by the general population. In the period of manic excitement, the mood elevation can become so intense that it can result in extended insomnia, extreme irritability, and heightened aggressiveness. In the period of depression, which may last for several weeks or months, a patient experiences feelings of general fatigue, uselessness, and hopelessness and, in serious cases, may contemplate suicide.

Text 10

Chinese Beauty

Under attack from some US church groups over the controversial film "Priest", The Walt Disney Co. is no longer to protest. This time they have provoked the anger of a government. China's leaders are strenuously objecting to Disney's decision to distribute "Kundun", currently being produced under the direction of Martin Scorsese. The movie tells the story of Tibet's exiled spiritual leader, the Dalai Lama. China's leaders are the villains of the piece and are not amused, going so far as to suggest that Disney's future share of the enormous potential media market in China may be under threat if they go ahead with their plans to distribute the movie.

Last year a number of major Hollywood studios are reported to have turned down the chance to distribute "Kundun" for fear of upsetting the Chinese.

Text 11

The extinction of many species of birds has undoubtedly been hastened by modern man; since 1600 it has been estimated that approximately 100 bird species have become extinct over the world. In North America, the first species known to be annihilated was the great auk, a flightless bird that served as an easy source of food and bait for Atlantic fishermen through the beginning of the nineteenth century.

Shortly after the great auk’s extinction, two other North American species, the Carolina parakeet and the passenger pigeon, began dwindling noticeably in numbers. The last Carolina parakeet and the last passenger pigeon in captivity both died in September 1914. In addition to these extinct species, several others such as the bald eagle, the peregrine falcon, and the California condor are today recognized as endangered; steps are being taken to prevent their extinction.

Text 12

You may want to gain a position of greater responsibility in your current field. Or you may want to switch from one area of business to another - from marketing or sales to strategic planning or finance, for example. An MBA can give you '.he added clout you need to get that job. Using an MBA to switch career focus is a typical strategy.

Many people seek a management position after acquiring expertise in a different field altogether such as nursing, teaching, performing arts, or engineering because they want to shift to a management position in their area. An MBA can help you make a major break or leap in your career path - with a concomitant leap in income. Or you may aspire to an area that "requires" an MBA as an entry. An MBA signals a prospective employer that you've got the knowledge and skills a demanding job requires, as well as persistence, energy, and even time-management skills to handle it.

Text 13

While most desert animals will drink water if confronted with it, for many of them the opportunity never comes. Yet all living things must have water, or they will expire. The herbivores or animals that eat only plants find it in desert plants. The carnivores, which eat only meat, slake their thirst with the flesh and blood of living prey. One of the most remarkable adjustments, however, has been made by the tiny kangaroo rat, which not only lives without drinking but subsists on a diet of dry seeds containing about 5% free water. Like other animals, it has the ability to manufacture water in his body by a metabolic conversion of carbohydrates. But it is notable for the parsimony with which it conserves its small supply by every possible means, expanding only minuscule amounts in his excreta and through evaporation from his respiratory tract.

Text 14

Bighorn sheep eat the grass and other vegetation that flourishes in the high mountain basins and valleys of the Rocky Mountains. Usually these places lie between the snow line and a point a little below the timber line. This is also the area where the elderly rams fight their thunderous battles for leadership of the flocks during the mating season.

In preparation for these brutal encounters, they select a piece of level ground where they can face each other at a distance of ten to fifteen feet. Suddenly and at the same instant, both rams charge forward with lowered heads, their huge, curved horns meeting with a crash that would rival a freeway collision. Often the force of the collision throws both battlers back on their hind legs. But they keep on charging and crashing until one or the other decides that he has been beaten.

Perhaps you wonder why neither ram breaks its neck or its horns in these wild battles. The spine is protected by a special, extremely strong rubber-like reinforcement at the danger spot where it joins the skull. The horns are safe because the blow is always struck with their thickest central part, their thinner curving points being entirely outside the danger zone.

Once leadership has been established, peace returns to the battlefield, and the bighorn sheep go about their business.

Text 15

Robot Birds

Liverpool city council want to clear the city of fat pigeons. They say that people are feeding the birds, which makes them fat. The pigeons get bigger because their normal diet would consist of seeds and insects, not high-fat junk food they are eating in the city centre.

The council want people to know that everyone who feeds the pigeons is responsible for the streets being so crowded with these birds. They hope to encourage the birds to move away from the city centre and into parks and open spaces.

Ten robotic birds have been brought into the city centre to scare the pigeons away and visitors are asked not to give the pigeons any food. The mechanical birds – known as 'robops' – will sit on the roofs of buildings. They can be moved around to different locations. They look like a peregrine falcon, which is a bird that kills pigeons. They even make noises and flap their wings to scare the pigeons. They hope that the pigeons will go away before the city becomes the European Capital of culture in two years.

Text 16

New studies on air quality inside office buildings show that the indoor air quality is more hazardous to human health than the polluted air outside. Each year, the air in our cities exceeds safe levels during at least 60 days of the summer. According to the Committee on the Environment, the air quality in approximately 30% of buildings seems unsafe.

Medical conditions, including asthma, cancer, and depression, may be connected to poor indoor air quality. Cleaning product, furniture, air conditioners, and gas heating systems all contribute to poor indoor air quality. The most common reason for Sick Building Syndrome, a medical condition, has been blamed on poor indoor air quality. When a building opens too early, paint fumes and cleaning products do not have enough time to disperse. These fumes can remain in the air for a long time. They can affect customers or clients, and particularly building staff.

Text 17

Intelligence pills

Some scientists have predicted that healthy adults and children may one day take drugs to improve their intelligence and intellectual performance. A research group has suggested that such drugs might become as common as coffee or tea within the next couple of decades.

To counter this, students taking exams might have to take drugs tests like athletes. There are already drugs that are known to improve mental performance, like Ritalin, which is given to children with problems concentrating. A drug given to people with trouble sleeping also helps people remember numbers.

These drugs raise serious legal and moral questions, but people already take vitamins to help them remember things better, so it will not be a simple problem to solve. It will probably be very difficult to decide at what point a food supplement becomes an unfair drug in an examination.

Only children will take pills to improve their intellectual performance.

Text 18

Busy people do not want their vacations to be a hassle. That's why all-inclusive resorts are becoming popular. At these resorts, one price includes all meals, drinks, lodging, and sightseeing. Golf, tennis, and swimming are available for free. Other sports, such as scuba diving, deep sea fishing, and rock climbing, may require separate fees for equipment rental, but instruction and excursions are included. Many resorts also include children's activities as part of the package. Check with a travel agent to find an all-inclusive resort with activities you would enjoy.

Text 19

Is your company a sitting duck for hackers? When did you last change your password? How complete are your security systems? Have you ever been broken into before?

According to IANS, the International Association for Network Security, there is a new breed of hackers out there. And, there is a new target.

In the past, hackers gained notoriety from breaking into big companies' networks. In fact, the bigger the company, the bigger the success. When hackers broke into Infelmax's notoriously secure system in 1999, they made headline news around the world.

The big "successes" came with a major drawback. These headline break-ins came with international teams of investigators and serous criminal charges. Several former hackers are now sitting behind bars or working overtime to pay off hefty fines in penalties and damages.

So, hackers of the new decade have turned to a new target: smaller companies. Smaller companies often spend less on their security systems. If they have never been broken into before, they may be lulled into a feeling of security. They are often lax about changing their password frequently enough. And that spells trouble. However, if you do fall victim to hackers, it will definitely attract your attention. These thieves can gain access to your files, destroying, copying, or alerting them. They can create havoc with your data. And if they do, you'll surely wish you had changed your password more often.

Text 20

Notebook PCs are gaining popularity because of their convenience. They are lighter and smaller than desktop computers and much easier to carry around. Although laptop and notebook computers are more expensive than desktop computers, more and more people are buying them. They are filling a growing need for mobility. The trend towards giving electronic items as holiday gifts is also growing. The old-fashioned approach to holiday celebrations is giving way to the enthusiasm for new technology.

Text 21

The advent of new technology is threatening to shut down a major sector of the economy. Several new discoveries made by Bowing Oil have allowed some manufacturers to weave cloth made from petroleum products. These new fabrics are unlike rayons, polyesters, and acetates. They are more like cotton, hemp, and wool because they are cool and breathe.

These new fabrics were developed over a period of twenty years and along the way scientists encountered many problems, including disintegration, bleeding and fading. However, new manufacturing techniques overcame such setbacks and have eventually allowed for more natural threads to be produced.

In some cases, it is impossible to tell the difference between the natural and the new fabrics. These new fabrics are cheaper to produce than cotton and as a result, some mills are not buying from cotton or wool producers. In one region near Because, over ninety percent of the cotton producers have shut down their operations.

Text 22

Forecasts of storms

A prominent forecast predicts that 11 storms, including 7 hurricanes, will form over the Atlantic ocean this year.

Professor Julian Ross of the National Meteorological Center has been predicting the number and intensity of hurricane season storms for a dozen years. His latest report is an update on an earlier forecast where he predicted a below-average season with eight storms, five of them hurricanes.

Although not always on target, the Professor's predictions are more accurate than most. Last year he predicted 18 storms, with 10 hurricanes, including two major ones. There were in fact 18 storms, but 14 of these were hurricanes and 4 of them were major. It was one of the busiest hurricanes seasons on record. Forty-five people lost their lives, and damages in the southern Atlantic states was estimated in the billions of dollars.

Text 23

Too much Canadian Wood?

American timber producers have announced that unless Canada can be persuaded to limit exports of softwood to the United States, they will ask that import duties be imposed. They claim that subsidized Canadian timber has taken 36% of the American market and that 29,000 jobs were lost last year as a result. North America produced a glut of timber in 1995 as mills worked flat out to make wood chips (used in paper-making) when paper prices soared. But paper prices have since collapsed and mills are shutting or cutting output. Demand for timber is rising again in Japan, which should divert some Canadian wood.

Text 24

Puppy love may indeed feel like its deeper emotional cousin called "true love," but it usually has a fleeting quality which keeps it from developing any further than a serious crush. Puppy love often starts with a one-sided infatuation, perhaps for a seemingly unapproachable classmate or an attractive teacher or other authority figure. For the smitten one, these feelings of attraction may be very intense indeed, and create the sort of distractions and obsessive thought patterns of true love.

Some young people gripped by puppy love will eventually make an effort to pursue a relationship with the object of their affection, with varying degrees of success. Occasionally two people with similar afflictions will embark on a short adolescent romance with little to no expectations of a long-term commitment. Others may lack the social confidence to approach the object of their affections, keeping their feelings at the level of an unrequited crush.

Puppy love rarely reaches a level of concern for parents, although they may become concerned if their child's budding romance begins to interfere with their other obligations and basic needs. A young adolescent experiencing puppy love for the first time may become easily distracted or less focused on other matters such as eating, sleeping and schoolwork. The inevitable break-up of a puppy love relationship may also create deep emotions which a young teen may not be prepared to handle. Parents should respect the fact that a puppy love relationship can feel very real indeed to the participants, and often mimics the same stages as a more mature romantic relationship.

The concept of a youthful infatuation or immature adolescent romance can be traced back for centuries. During the time of Shakespeare, the idea of puppy love was often rendered as calf love, with the same basic premise of an immature romantic crush. In Romeo and Juliet, Romeo is reminded of his former "calf love," a girl named Rosaline. Romeo's budding relationship with Juliet could be construed as a transition from a fleeting puppy love into a true, if tragic, romance for the ages.

Text 25

Nurture and Teach: The single most important thing caregivers can do for a child is provide a nurturing environment. By doing this, we influence children's brain development and their ability to learn. Introducing nurtured children to learning opportunities every day will help them become happy, well-adjusted adults. In all stages of child development, each experience builds on the one before it. The most basic foundations can serve as the basis for the comprehension of more complex ideas in future years.

We are born with billions of brain cells; in fact, all we will ever have. What is missing is synapses, a large amount of connections between those brain cells. They start developing based on a child's experiences. Children's brains develop faster from birth to age three than any other time; and more learning takes place during this time than any other. The more learning opportunities parents provide for their children from birth until school age, the more synapses are made. The connections will serve as a pool of knowledge for a child to access in later years.

Not only do children need to be physically active, it is their nature to look for opportunities to learn. They participate in learning by using their senses and asking countless questions in order to more fully understand the task at hand. Children enjoy learning when they can master an activity. Begin with a simple task and expand or complicate it after your child has enjoyed some successes. Creating a safe and secure learning environment will help children do their best learning rather than distracting them. Key to creating this environment is treating your child with respect and caring.

The main way children collect and process information is through play. Play is the repetition that reinforces old skills and encourages new ones. Because play is enjoyable, children's minds are opened. Children are capable of much learning through play because they are very receptive and relaxed. Take advantage of this benefit and select activities that are fun and educational. Your children will learn, and you will both enjoy the experience.

Text 26

Propaganda and the media

What is propaganda?

Politicians want to win our hearts and minds, and businesses want to win our dollars. Both these groups use the media to manipulate public opinion and to try to gain mass support. This deliberate manipulation of public opinion is called “propaganda.” It may be true or false, but what sets propaganda apart from other communication is that it is intended to shape our views or change our opinions.

How does propaganda work?

A number of methods of propaganda are used in the media. One is giving positive or negative labels. If something is given a negative label, then the audience might reject the idea or person without questioning the label. Likewise, a positive label might encourage people to support an idea. For example, a public rally may be described in the media as “peaceful” or as “disruptive.” The different labels will give different impressions. People who favor limits on gun possession may be described as "voices for gun control" or "anti-gun activists." Some labels encourage a favorable response and others a negative one.

We can also be influenced by visual images in the media. A newspaper editor may have several photographs of a famous person to choose from. An attractive picture may be chosen to create a good impression, or an unattractive one to have the opposite effect.

Another method of propaganda is to have a beautiful or famous person promote some idea or product, or link the product to other concepts that are admired or desired. At election time, politicians often look for support from movie stars or famous sports people to campaign for them. Similarly, in advertising, products are promoted by beautiful, expensively dressed people. The audience is encouraged to make an association from the attractiveness of one image to the other.

Another method used by promoters is to create the impression that everyone supports the product or idea. Soft-drink companies use commercials to show a crowd of young, happy people all drinking their product. This is propaganda because it puts pressure on the audience to conform to this behavior because it is seen as popular.

Why doesn't propaganda always work?

There are limits on the role of propaganda. The propagandist, a person who produces propaganda, cannot simply make us believe in something, or easily get rid of beliefs or opinions we already hold. There are at least three factors that limit the influence of the media on public opinion. One is the fact that independent organizations can present us with different points of view, so the influence of one can cancel out or balance the influence of another. A second is that media owners are interested in making a profit, so the media often presents what the audience already wants to see or hear. The third is that there is often a two-step process of influence: we may hear an analysis of an issue on television, but we often accept or reject it after being influenced by other opinion leaders – people in the community whom we respect.

Text 27

Hay fever is a seasonal allergy to pollens; the term “hay fever,” however, is a less than adequate description since an attack of this allergy does not incur fever and since such an attack can be brought on by sources other than hay-producing grasses. Hay fever is generally caused by air-borne pollens, particularly ragweed pollen. The amount of pollen in the air is largely dependent on geographical location, weather, and season. In the eastern section of the United States, for example, there are generally three periods when pollen from various sources can cause intense hay fever suffering: in the springtime months of March and April when pollen from trees is prevalent, in the summer months of June and July when grass pollen fills the air, and at the end of August when ragweed pollen is at its most concentrated levels.

Text 28

Tea for Health

Drinking tea every day can help maintain your health

Tea for health is dedicated to bringing you the facts about how tea, the world’s favourite drink, can help you maintain your health as part of a balanced diet and healthy, active lifestyle. If you need questions answering about tea and looking after your health, or if you want to know why drinking at least 4 cups of tea a day is a good way for you to help maintain your health, then look no further – this site provides a wealth of information that can help you.

For the science behind the facts, health professionals can access peer-reviewed evidence in a bespoke section of the site. The latest research from around the world is presented and you can also register for the ‘updates’ service so you’ll see the most recent research, first. The site provides a number of Monographs to help you with the background that can be downloaded in PDF format, including an information leaflet that also gives an overview on why tea can be an important part of a healthy lifestyle.

Text 29

The rattlesnake has a reputation as a dangerous and deadly snake with a fierce hatred for humanity. Although it is indeed a venomous snake capable of killing a human, its nature has perhaps been somewhat exaggerated in myth and folklore.

It is not inherently aggressive and generally strikes only when it has been put on the defensive. In its defensive posture, it raises the front part of its body off the ground and assumes and S-shaped form in preparation for a lunge forward. At the end of a forward thrust, it pushes its fangs into the victim, thereby injecting its venom.

There are more than 30 species, varying in length from 20 inches to 6 feet and also varying in toxicity of venom. In the United States there are only a few deaths annually from this kind of snakes, with a mortality rate less than 2 percent of those attacked.

Text 30

No More Darkness

What do you do when power failures plunge your home into darkness? Some ingenious engineers and technicians from Auto Design Inc. have a simple solution: go outside and start your car. Company researchers discovered that for only a few hundred dollars, a car engine could be equipped to provide all the electrical power required by a house.

Auto Design has applied for a patent and hopes to unveil the engine-generator at the annual Automotive Engineers show in February. The simplicity of the design and the low cost of the necessary engine modifications could make this innovation very attractive to the general public. So the next time blizzards, hurricanes or other disasters interfere with electrical service, forget about searching in the dark for candles or flashlights. Instead, grab your car keys, start the engine and watch the darkness disappear.

APPENDIX Q

Readability Scores from Three Indications of Text Difficulty by LFP, Flesch Reading Ease Score and Flesch-Kincaid Grade Level

Text	Number of Frequency Bands Covering 95% of tokens	% of Tokens from the First Three Frequency Bands	% of Tokens from the First Frequency Band	Flesch Reading Ease Score	Flesch-Kincaid Grade Level
1	5	89.52	84.76	29.8	12.0
2	5	87.70	72.31	42.2	11.9
3	10	84.39	70.63	41.2	12.0
4	15	89.84	80.49	40.7	12.0
5	5	89.93	79.14	32.5	12.0
6	10	86.75	70.86	41.7	12.0
7	4	93.39	73.53	34.7	12.0
8	6	87.33	66.90	18.9	12.0
9	7	88.18	64.55	15.3	12.0
10	15	85.11	69.50	47.1	12.0
11	13	83.34	68.12	32.4	12.0
12	5	92.59	79.63	50.3	11.2
13	8	86.12	74.31	45.7	12.0
14	5	88.00	73.33	60.4	9.9
15	4	94.57	78.26	64.5	8.3
16	3	96.59	85.71	46.9	10.8
17	3	96.08	79.74	50.9	10.8
18	9	93.41	71.43	45.6	10.0
19	7	89.36	78.30	51.6	8.8
20	11	89.47	77.63	34.7	11.5
21	6	89.74	75.00	43.3	11.5
22	6	87.32	66.42	41.8	11.6
23	4	93.33	73.33	47.5	11.0
24	8	84.35	73.91	35.9	12.0
25	3	95.71	88.54	52.3	9.3
26	13	83.34	68.12	43.0	11.2
27	9	84.28	65.71	42.4	12.0
28	4	94.63	88.71	50.4	12.0
29	9	88.36	73.29	52.5	11.2
30	6	90.08	76.60	38.4	12.0

APPENDIX R

Predicted Reading Comprehension Scores at Different Vocabulary Size

LFP1

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
50	0.06	0
60	0.12	0
70	0.19	0
80	0.25	0
90	0.32	0
100	0.38	0
110	0.45	0
120	0.51	1
130	0.58	1
140	0.64	1
150	0.71	1
160	0.77	1
170	0.84	1
180	0.90	1
190	0.97	1
200	1.03	1
210	1.10	1
220	1.16	1
230	1.23	1
240	1.29	1
250	1.36	1
260	1.42	1
270	1.49	1
280	1.55	2
290	1.62	2
300	1.68	2
310	1.75	2
320	1.81	2
330	1.88	2
340	1.94	2
350	2.01	2

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
360	2.07	2
370	2.14	2
380	2.20	2
390	2.27	2
400	2.33	2
410	2.40	2
420	2.46	2
430	2.53	3
440	2.59	3
450	2.66	3
460	2.72	3
470	2.79	3
480	2.85	3
490	2.92	3
500	2.98	3
510	3.05	3
520	3.11	3
530	3.18	3
540	3.24	3
550	3.31	3
560	3.37	3
570	3.44	3
580	3.50	4
590	3.57	4
600	3.63	4
610	3.70	4
620	3.76	4
630	3.83	4
640	3.89	4
650	3.96	4
660	4.02	4
670	4.09	4
680	4.15	4
690	4.22	4
700	4.28	4
710	4.35	4
720	4.41	4
730	4.48	4
740	4.54	5
750	4.61	5
760	4.67	5
770	4.74	5
780	4.80	5

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
790	4.87	5
800	4.93	5
810	5.00	5
820	5.06	5
830	5.13	5
840	5.19	5
850	5.26	5
860	5.32	5
870	5.39	5
890	5.52	6
900	5.58	6
910	5.65	6
920	5.71	6
930	5.78	6
940	5.84	6
950	5.91	6
960	5.97	6
970	6.04	6
980	6.10	6
990	6.17	6
1000	6.23	6

LFP 2

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
90	0.01	0
100	0.08	0
110	0.14	0
120	0.21	0
130	0.28	0
140	0.35	0
150	0.42	0
160	0.49	0
170	0.56	1
180	0.63	1
190	0.70	1
200	0.76	1
210	0.83	1
220	0.90	1
230	0.97	1
240	1.04	1

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
250	1.11	1
260	1.18	1
270	1.25	1
280	1.32	1
290	1.38	1
300	1.45	1
310	1.52	2
320	1.59	2
330	1.66	2
340	1.73	2
350	1.80	2
360	1.87	2
370	1.94	2
380	2.01	2
390	2.07	2
400	2.14	2
410	2.21	2
420	2.28	2
430	2.35	2
440	2.42	2
450	2.49	2
460	2.56	3
470	2.63	3
480	2.69	3
490	2.76	3
500	2.83	3
510	2.90	3
520	2.97	3
530	3.04	3
540	3.11	3
550	3.18	3
560	3.25	3
570	3.32	3
580	3.38	3
590	3.45	3
600	3.52	4
610	3.59	4
620	3.66	4
630	3.73	4
640	3.80	4
650	3.87	4
660	3.94	4
670	4.00	4
680	4.07	4

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
690	4.14	4
700	4.21	4
710	4.28	4
720	4.35	4
730	4.42	4
740	4.49	4
750	4.56	5
760	4.62	5
770	4.69	5
780	4.76	5
790	4.83	5
800	4.90	5
810	4.97	5
820	5.04	5
830	5.11	5
840	5.18	5
850	5.25	5
860	5.31	5
870	5.38	5
880	5.45	5
890	5.52	6
900	5.59	6
910	5.66	6
920	5.73	6
930	5.80	6
940	5.87	6
950	5.93	6
960	6.00	6
970	6.07	6
980	6.14	6
990	6.21	6
1000	6.28	6

LFP3

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
50	0.01	0
60	0.08	0
700	0.14	0
80	0.20	0
90	0.26	0

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
100	0.32	0
110	0.38	0
120	0.44	0
130	0.51	1
140	0.57	1
150	0.63	1
160	0.69	1
170	0.75	1
180	0.81	1
190	0.87	1
200	0.94	1
210	1.00	1
220	1.06	1
230	1.12	1
240	1.18	1
250	1.24	1
260	1.31	1
270	1.37	1
280	1.43	1
290	1.49	1
300	1.55	2
310	1.61	2
320	1.67	2
330	1.74	2
340	1.80	2
350	1.86	2
360	1.92	2
370	1.98	2
380	2.04	2
390	2.10	2
400	2.17	2
410	2.23	2
420	2.29	2
430	2.35	2
440	2.41	2
450	2.47	2
460	2.54	3
470	2.60	3
480	2.66	3
490	2.72	3
500	2.78	3
510	2.84	3
520	2.90	3
530	2.97	3

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
540	3.03	3
550	3.09	3
560	3.15	3
570	3.21	3
580	3.27	3
590	3.33	3
600	3.40	3
610	3.46	3
620	3.52	4
630	3.58	4
640	3.64	4
650	3.70	4
660	3.76	4
670	3.83	4
680	3.89	4
690	3.95	4
700	4.01	4
710	4.07	4
720	4.13	4
730	4.20	4
740	4.26	4
750	4.32	4
760	4.38	4
770	4.44	4
780	4.50	5
790	4.56	5
800	4.63	5
810	4.69	5
820	4.75	5
830	4.81	5
840	4.87	5
850	4.93	5
860	5.00	5
870	5.06	5
880	5.12	5
890	5.18	5
900	5.24	5
910	5.30	5
920	5.36	5
930	5.43	5
940	5.49	5
950	5.55	6
960	5.61	6
970	5.67	6

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
980	5.73	6
990	5.79	6
1000	5.86	6

LFP4

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
0	0.98	1
10	1.02	1
20	1.06	1
30	1.10	1
40	1.13	1
50	1.17	1
60	1.20	1
70	1.24	1
80	1.28	1
90	1.32	1
100	1.35	1
110	1.39	1
120	1.43	1
130	1.46	1
140	1.50	2
150	1.54	2
160	1.57	2
170	1.61	2
180	1.65	2
190	1.69	2
200	1.72	2
210	1.76	2
220	1.80	2
230	1.83	2
240	1.87	2
250	1.90	2
260	1.95	2
270	1.98	2
280	2.02	2
290	2.06	2
300	2.09	2
310	2.13	2
320	2.17	2
330	2.20	2

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
340	2.24	2
350	2.28	2
360	2.32	2
370	2.35	2
380	2.39	2
390	2.43	2
400	2.46	2
410	2.50	3
420	2.54	3
430	2.58	3
440	2.61	3
450	2.65	3
460	2.69	3
470	2.72	3
480	2.76	3
490	2.80	3
500	2.84	3
510	2.87	3
520	2.91	3
530	2.95	3
540	2.98	3
550	3.02	3
560	3.06	3
570	3.09	3
580	3.13	3
590	3.17	3
600	3.21	3
610	3.24	3
620	3.28	3
630	3.32	3
640	3.35	3
650	3.39	3
660	3.43	3
670	3.47	3
680	3.50	4
690	3.54	4
700	3.58	4
710	3.61	4
720	3.65	4
730	3.69	4
740	3.72	4
750	3.76	4
760	3.80	4
770	3.84	4

Vocabulary Size Scores	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
780	3.87	4
790	3.91	4
800	3.95	4
810	3.98	4
820	4.02	4
830	4.06	4
840	4.10	4
850	4.13	4
860	4.17	4
870	4.21	4
880	4.24	4
890	4.28	4
900	4.32	4
910	4.35	4
920	4.39	4
930	4.43	4
940	4.47	4
950	4.50	5
960	4.54	5
970	4.58	5
980	4.61	5
990	4.65	5
1000	4.69	5

APPENDIX T

Predicted Reading Comprehension Scores at Different Text-specific Vocabulary Knowledge

TSVK	Predicted Reading Comprehension Scores	Round Predicted Reading Comprehension Scores
91	-0.21	0
92	0.42	0
93	1.06	1
94	1.69	2
95	2.33	2
96	2.96	3
97	3.60	4
98	4.23	4
99	4.87	5
100	5.50	6

Note: TSVK at 0-90% is not included in the table for two reasons: 1) the participants' minimum TSVK is about 92% and 2) when the TSVK is lower than 92%, the predicted reading comprehension score is negative and negative reading scores are impossible in this study.

APPENDIX U

Students' Rankings of Text Difficulty in the Questionnaire for Investigating Concurrent Validity of LFP as an Indicator of Text Difficulty

	Passage One	Passage Two	Passage Three	Passage Four
Student 1	1	3	2	4
Student 2	1	4	3	2
Student 3	1	2	3	4
Student 4	1	3	2	4
Student 5	2	1	3	4
Student 6	1	3	4	2
Student 7	1	2	3	4
Student 8	1	2	4	3
Student 9	2	1	3	4
Student 10	1	2	4	3
Student 11	1	3	2	4
Student 12	1	2	4	3
Student 13	1	2	3	4
Student 14	1	2	3	4
Student 15	1	3	2	4
Student 16	1	2	3	4
Student 17	1	2	4	3
Student 18	1	3	2	4
Student 19	1	2	4	3
Student 20	1	3	2	4
Student 21	1	2	3	4
Student 22	1	2	4	3
Student 23	1	3	2	4
Student 24	1	2	4	3
Student 25	1	3	2	4
Student 26	1	2	3	4
Student 27	1	2	4	3
Student 28	1	2	4	3
Student 29	1	3	2	4
Student 30	1	2	4	3
Student 31	1	2	3	4
Student 32	1	3	2	4

	Passage One	Passage Two	Passage Three	Passage Four
Student 33	1	2	4	3
Student 34	1	2	3	4
Student 35	1	2	3	4
Student 36	1	2	3	4
Student 37	1	3	2	4
Student 38	1	2	4	3
Student 39	1	3	2	4
Student 40	1	3	2	4
Student 41	1	2	4	3
Student 42	1	2	4	3
Student 43	1	3	2	4
Student 44	1	3	2	4
Student 45	1	3	2	4

APPENDIX V

Teachers' Rankings of Text Difficulty in the Questionnaire

for Investigating Concurrent Validity of LFP

as an Indicator of Text Difficulty

	Passage One	Passage Two	Passage Three	Passage Four
Teacher 1	1	2	3	4
Teacher 2	1	2	4	3
Teacher 3	1	3	2	4
Teacher 4	1	3	2	4
Teacher 5	1	2	4	3

APPENDIX W

Readability Scores from Flesch Readability Formula

and Flesch Readability Formula

	Passage One	Passage Two	Passage Three	Passage Four
Flesch Readability Formula	71.0	63.7	50.6	38.2
Flesch-Kincaid Readability Test	7.8	8.5	10.8	12.0

Students' Ratings of Text Difficulty in the Questionnaire for Investigating Predictive Validity of the Prediction of Reading Comprehension

Students	Passage One	Passage Two	Passage Three	Passage Four
A	5	4	5	3
B	4	4	4	2
C	5	4	3	2
D	5	4	3	3
E	4	3	3	2
F	4	3	3	3
G	4	3	3	1
H	5	4	4	3
I	3	3	3	3
J	4	3	3	2
K	3	3	3	2
L	3	3	3	1
M	3	2	3	2
N	4	3	3	3
O	4	3	3	1
P	4	2	3	2
Q	4	3	2	1
R	5	4	4	3
S	5	3	3	2
T	5	5	4	3
U	4	4	3	2
V	4	3	3	3
W	3	3	3	2
X	3	2	2	1
Y	4	4	3	2
Z	3	2	2	1
AA	3	3	2	1
AB	4	3	3	3
AC	5	4	4	2
AD	4	4	3	3

APPENDIX Y

Predicted Levels of Text Difficulty by Vocabulary Size

Students	Passage One	Passage Two	Passage Three	Passage Four
A	5	4	4	4
B	5	4	3	2
C	5	5	4	3
D	5	4	3	2
E	4	3	2	1
F	4	3	3	2
G	4	2	2	1
H	5	4	4	2
I	3	3	3	2
J	3	3	3	2
K	3	3	3	2
L	4	4	3	3
M	4	2	2	1
N	4	4	3	3
O	4	2	2	1
P	3	3	2	2
Q	3	3	2	1
R	5	3	3	2
S	4	3	2	1
T	5	4	4	3
U	4	4	3	3
V	4	3	3	2
W	3	3	2	2
X	4	2	2	1
Y	3	3	2	2
Z	4	1	1	1
AA	2	3	2	1
AB	4	4	3	3
AC	5	4	4	3
AD	5	4	3	2

APPENDIX Z

Predicted Levels of Text Difficulty by Text-specific Vocabulary Knowledge

Students	Passage One	Passage Two	Passage Three	Passage Four
A	5	5	5	5
B	4	5	4	4
C	5	5	4	4
D	5	5	4	4
E	5	4	4	4
F	5	5	3	4
G	5	4	4	3
H	5	5	5	5
I	5	4	4	5
J	5	5	3	4
K	4	5	4	4
L	5	5	4	3
M	4	4	3	4
N	5	5	4	5
O	5	4	4	3
P	5	4	4	4
Q	5	4	4	4
R	5	5	5	5
S	5	5	4	4
T	5	5	5	5
U	5	5	3	4
V	5	5	4	5
W	4	4	4	4
X	4	3	3	4
Y	5	5	5	4
Z	4	4	3	3
AA	5	5	4	3
AB	5	5	4	4
AC	5	5	4	5
AD	5	5	4	4

CURRICULUM VITAE

Ms. Patteera Thienpermpool was born on July 31, 1979. She graduated from Kasetsart University with a B.A. in English (2^{nd} class honours) in 1999. She obtained an M.A. in Applied Linguistics (English Language Teaching) from King Mongkut's University of Technology Thonburi in 2003. She has been working as a lecturer at the Faculty of Education, Silpakorn University since 2003. She undertook the co-supervision programme between the School of English, Institute of Social Technology, Suranaree University of Technology, Thailand and the Department of Linguistics, University of Essex, United Kingdom for the Degree of Doctor of Philosophy in English Language Studies. Her research interests include vocabulary, readability and language testing.