

**A COMPARISON OF ENGLISH READING ACHIEVEMENT USING THE
COMPUTER AND CONVENTIONAL MODES
OF TESTING**



Miss Sudthida Pavavijarn

สถาบันวิทยบริการ
**A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in English as an International Language (Inter-Department)**

Graduate School

Chulalongkorn University

Academic Year 2004

ISBN 974-17-6582-7

การเปรียบเทียบสัมฤทธิผลการอ่านภาษาอังกฤษด้วยวิธีการทดสอบ

แบบใช้เครื่องคอมพิวเตอร์และแบบประเพณีนิยม



นางสาวสุดธิดา ภาวะวิจารณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต

สาขาวิชา ภาษาอังกฤษเป็นภาษานานาชาติ (สหสาขาวิชา)

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2547

ISBN 974-17-6582-7

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title A Comparison of English Reading Achievement Using the
 Computer and Conventional Modes of Testing
By Miss Sudthida Pavavijarn
Field of study English as an International Language
Thesis Advisor Professor Kanchana Prapphal, Ph.D.

Accepted by the Graduate School, Chulalongkorn University in Partial
Fulfillment of the Requirements for the Master's Degree

..... Dean of the Graduate School
(Assistant Professor M.R. Kalaya Tingsabadh, Ph.D.)

THESIS COMMITTEE

..... Chairman
(Associate Professor Suphat Sukamolson, Ph.D.)

..... Thesis Advisor
(Professor Kanchana Prapphal, Ph.D.)

..... Member
(Assistant Professor Apasara Chinwonno, Ph.D.)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

นางสาวสุดธิดา ภาวะวิจารณ์: การเปรียบเทียบสัมฤทธิ์ผลการอ่านภาษาอังกฤษด้วยวิธีการ
ทดสอบแบบใช้เครื่องคอมพิวเตอร์และแบบประเพณีนิยม (A COMPARISON OF ENGLISH
READING ACHIEVEMENT USING THE COMPUTER AND CONVENTIONAL
MODES OF TESTING) อ. ที่ปรึกษา: ศ.ดร. กาญจนา ปรามพาล จำนวนหน้า 167 หน้า
ISBN 974-17-6582-7

การวิจัยนี้มีวัตถุประสงค์เพื่อ (1) เปรียบเทียบสัมฤทธิ์ผลการอ่านเพื่อความเข้าใจในภาษาอังกฤษใน
ระดับมหาวิทยาลัยด้วยวิธีการทดสอบแบบใช้เครื่องคอมพิวเตอร์และแบบประเพณีนิยม และ (2) เปรียบเทียบ
เจตคติของผู้รับการทดสอบที่ศึกษาอยู่ในระดับมหาวิทยาลัยต่อวิธีการทดสอบแบบใช้เครื่องคอมพิวเตอร์และ
แบบประเพณีนิยม กลุ่มตัวอย่างเป็นนักศึกษาคณะมนุษยศาสตร์ มหาวิทยาลัยเชียงใหม่ ชั้นปีที่ 1 ซึ่งศึกษาใน
กระบวนวิชาภาษาอังกฤษพื้นฐาน 1, 2 และ 3 ในภาคการศึกษาที่ 1 ปีการศึกษา 2547 ที่มหาวิทยาลัยเชียงใหม่
จำนวน 66 คน ซึ่งแบ่งเป็น 3 กลุ่ม โดยใช้ผลคะแนนการสอบวิชาภาษาอังกฤษจากการสอบเข้าศึกษาต่อใน
ระดับมหาวิทยาลัย (Entrance Examination) เป็นเกณฑ์ในการแบ่งกลุ่ม เครื่องมือที่ใช้ในการวิจัยเป็นข้อสอบ
วัดความเข้าใจการอ่านภาษาอังกฤษจำนวน 2 ชุด นักศึกษาต้องทำข้อสอบทั้ง 2 ชุด ได้แก่ ชุดที่ 1 เป็นแบบ
ประเพณีนิยมซึ่งนักศึกษาตอบในกระดาษคำตอบ และชุดที่ 2 เป็นแบบใช้เครื่องคอมพิวเตอร์ซึ่งนักศึกษาตอบ
บนหน้าจอคอมพิวเตอร์ สำหรับแบบสอบถามเจตคติต่อการทดสอบด้วยทั้ง 2 วิธี นักศึกษาตอบก่อนและหลัง
การทดสอบ การวิเคราะห์ข้อมูลใช้ t-test เปรียบเทียบความแตกต่างของค่าเฉลี่ยของคะแนนจากการทดสอบ
ทั้ง 2 ชุด และค่าเฉลี่ยของคะแนนจากแบบสอบถามก่อนและหลังการทดสอบ ผลการวิจัยพบว่า (1) สัมฤทธิ์
ผลการอ่านความเข้าใจภาษาอังกฤษด้วยวิธีการทดสอบโดยใช้เครื่องคอมพิวเตอร์และแบบประเพณีนิยมมี
ความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 (2) เจตคติของนักศึกษาที่มีต่อการทดสอบทั้ง 2 วิธี
ก่อนและหลังการทดสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยที่นักศึกษามีเจตคติที่ดีต่อการ
ทดสอบแบบใช้เครื่องคอมพิวเตอร์มากกว่าการทดสอบแบบประเพณีนิยม

สหสาขาวิชา ภาษาอังกฤษเป็นภาษานานาชาติ ลายมือชื่อนิติ
ปีการศึกษา 2547 ลายมือชื่ออาจารย์ที่ปรึกษา.....

4589498220: ENGLISH AS AN INTERNATIONAL LANGUAGE

KEY WORDS: ENGLISH READING ACHIEVEMENT, READING COMPREHENSION TESTS, COMPUTER-BASED TESTS, LANGUAGE TESTS AND TECHNOLOGY, ATTITUDES AND TESTING, ATTITUDES TOWARDS TECHNOLOGY IN TESTING, MODES OF TESTING

MISS SUDTHIDA PAVAVIJARN: A COMPARISON OF ENGLISH READING ACHIEVEMENT USING THE COMPUTER AND CONVENTIONAL MODES OF TESTING.

**THESIS ADVISOR: PROFESSOR KANCHANA PRAPPHAL, Ph.D.,
167 pp. ISBN 974-17-6582-7**

The study aimed to (1) compare reading comprehension in English using the computer and conventional modes of testing at the tertiary level and (2) compare university students' attitudes towards the two modes of testing. The sample was a group of 66 first year university students from the Faculty of Humanities at Chiang Mai University. They studied Foundation English (FE I, II, and III) courses during the first semester of the academic year of 2004 at Chiang Mai University. They were classified into three levels of English proficiency based on their English language test scores from the Entrance Examination. The instruments used in the investigation were of two types: (1) two sets of Reading Comprehension Test of English (RCTE) and (2) a questionnaire on students' attitudes towards the two modes of testing. The RCTE was delivered via computers and on printed test booklets. The students took both tests. The questionnaire was distributed before and after the administration of the tests. The data collected underwent statistical analyses. The mean scores from the two tests were compared using the t-test of dependent sample formula. The mean scores of the questionnaire were also compared using the t-test of the same formula. The results of the analyses revealed that (1) there was no significant difference between the mean scores from the computer-based test and that from the conventional test at the significant level of .05 and (2) there were significant differences between the mean scores of the attitude towards the computer-based test and that of the conventional test at the significant level of .05 before and after the administration of the tests. In other words, the students showed positive attitudes towards the computer-based tests in this study.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department English as an International Language Student's signature _____
Field of study -- Advisor's signature _____
Academic year 2004

ACKNOWLEDGEMENTS

Knowledge learned in classes and from books is to apply to reach other new and beneficial knowledge and contribute them to the society. In doing this kind of learning, I feel that I have been growing up as a responsible individual as well as a potential language teacher. Along this journey of learning I have never felt abandoned. I would like to take this opportunity to express my sincere gratitude to Professor Kanchana Prapphal, Ph.D., the advisor of this thesis who has introduced a course in language testing to us, and I truly appreciate her for dedicating hours in reading and giving suggestions to improve the thesis. Other people may mistakenly perceive language testing as complicated. Professor Kanchana has shown that it is not a life-threatening subject. I would also like to thank the thesis committee, Associate Professor Suphat Sukamolson, Ph.D. and Assistant Professor Apasara Chinwonno, Ph.D. for making this thesis a complete one. I would like to forward many thanks to five specialists, Assistant Professor Pavinee Thirakhupt, Ph.D., Assistant Professor Krongkeo Kannasoot, Assistant Professor Apinya Phakdeewongse, Associate Professor Arunee Wiriyachitra, and Associate Professor Varunee Boon-Long, Ph.D. who helped in validating the instruments used in this study.

Besides, I would like to express my deep appreciation to my parents who have supported me all along through many difficult times. I am deeply grateful to all of my friends and colleagues who have given long distance and consistent consolation and suggestions. I would like to forward many thanks to my foreign friends for proofreading the tests and this thesis. This study would not be possible without great assistance from a brilliant computer programmer, Khun Narong. Moreover, I sincerely thank all of the students who sacrificed hours to participate in this study though it was not directly related to their studies. Last but not least, I would like to extend thanks to three prospective English major students who helped in the data collection procedure and the three computer technicians at the Faculty of Humanities at Chiang Mai University who were there to make sure the computers worked smoothly.

Thank you.

CONTENTS

	Page
ABSTRACT IN THAI	iv
ABSTRACT IN ENGLISH	v
ACKNOWLEDGEMENTS	vi
CONTENTS	vii
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
CHAPTER I: INTRODUCTION	1
Background of the study	1
Statement of the problem	4
Purposes of the study	6
Research questions	7
Statement of hypotheses	7
Variables	7
Scope of the study	8
Limitations of the study	9
Definitions of terms	12
Significance of the study	14
Summary	16
CHAPTER II: REVIEW OF RELATED LITERATURE	17
Reading models	17
1. Bottom-up models.....	17
2. Top-down models.....	18
3. Interactive models.....	19
Reading in the first (L1) and the second/foreign language (SL/FL)	22
1. Processes of reading in L1 and SL/FL.....	22
1.1 Coady's psycholinguistic model.....	23

1.2 Bernhardt's constructivist model.....	23
1.3 Second/Foreign language reading and schema theory..	24
2. Distinctions between reading in the first and the second/foreign language.....	25
Reading comprehension	27
Reading skills in the SL/FL	29
Factors affecting reading in the SL/FL	33
1. Factors related to the reader.....	33
1.1 Roles of schemata.....	33
1.2 Reader motivation and affect.....	35
1.3 Beginning readers and fluent readers.....	35
2. Factors related to the text.....	36
2.1 Topic and content.....	36
2.2 Text type, genre, and organization.....	36
2.3 Traditional linguistic factors.....	37
2.4 Text reliability.....	37
2.5 Other related factors.....	37
Applications of computers in language testing	38
1. Advantages of the use of the computer in language testing	39
1.1 Benefits for test developers.....	39
1.2 Benefits for test takers.....	40
2. Disadvantages of the use of the computer in language testing...	41
3. Considerations in the application of the computer in language testing.....	41
3.1 Issues concerning test purposes and constructs.....	42
3.2 Issues concerning computer-adaptive/assisted language testing (CALT).....	42
4. Computerized test delivery systems.....	43
4.1 Computerized fixed tests.....	43

4.2 Automated test assembly for computer delivery.....	44
4.3 Computerized adaptive testing.....	44
4.4 Computerized classification tests.....	45
Studies related to comparing conventional and computerized tests....	46
1. Equivalence of the conventional and the computer-based tests.....	46
2. Studies on mode effects.....	49
3. Students' attitudes towards the computer-based materials.....	51
Summary.....	53
CHAPTER III: METHODOLOGY AND PROCEDURES	54
Research design.....	54
Population and sample of the study.....	54
1. The population.....	54
2. Sample.....	55
Instruments of the study.....	57
1. Development of the Reading Comprehension Test of English (RCTE).....	57
1.1 Defining the test constructs.....	57
1.2 Design statement of the RCTE.....	58
1.3 The test blueprint.....	63
2. The development of the questionnaire.....	69
2.1 Purpose.....	69
2.2 Constructs and organization.....	69
3. The development of the computer program to deliver the RCTE.....	70
Procedures.....	73
1. Development of the instruments.....	73

2. The instruments validation.....	73
2.1 A priori validation.....	74
2.1.1 Qualitative analysis of the RCTE.....	74
2.1.2 Qualitative analysis of the questionnaire.....	75
2.2 A posteriori validation.....	76
2.2.1 The RCTE item analysis.....	76
2.2.2 Correlation between the pilot paper-based RCTE, the computer-based RCTE, and the English test scores from the Entrance Examination.....	78
2.2.3 The RCTE items selection.....	80
2.2.4 The quantitative analysis of the questionnaire.....	82
The data collection and data analyses.....	86
1. Data collection procedure and design.....	86
2. Data analyses.....	87
2.1 The quantitative analysis of the RCTE.....	87
2.2 The quantitative analysis of the questionnaire.....	90
Summary.....	93
CHAPTER IV: FINDINGS.....	94
Summary of the instruments validation.....	94
Hypotheses testing.....	94
1. Hypothesis 1 testing.....	95
2. Hypothesis 2 testing.....	96
Findings on the questionnaire.....	98
1. Interpretation of Part 1: students' familiarity with the computer.....	99

2. Interpretation of Part 2: students' attitudes towards the conventional and computerized RCTE.....	99
3. Findings from Part 3 of the questionnaire.....	102
Summary	103
CHAPTER V: DISCUSSION AND RECOMMENDATIONS	104
Summary of the study	104
1. Purposes of the study.....	104
2. Participants of the study.....	104
3. The instruments.....	105
4. Procedures of the study.....	105
5. Data collection and analyses.....	106
Summary of findings	106
1. Findings on the test performances.....	106
2. Findings on the attitudes towards the two modes of testing.....	107
Discussion of the findings	108
1. The RCTE results.....	108
1.1 Comparison of the test scores from the computerized and conventional RCTE.....	108
1.2 The test scores from the computerized and conventional RCTE.....	109
1.3 Relationships between the computerized RCTE, conventional RCTE, and the English scores from the Entrance Examination.....	110
2. The attitudes towards the computer and conventional modes of testing.....	111

2.1 Comparison of the mean scores from the attitude questionnaire.....	111
2.2 Positive attitudes towards the computerized test.....	112
2.3 Additional findings from the questionnaire.....	112
Conclusions	113
Implications for language assessment	113
Recommendations for future studies	115
REFERENCES	118
APPENDICES	128
Appendix A.....	129
Appendix B.....	140
Appendix C.....	145
Appendix D.....	147
Appendix E.....	154
Appendix F.....	158
Appendix G.....	160
Appendix H.....	164
BIOGRAPHY	168

LIST OF TABLES

		Page
Table 1.	Matrix of Reading Types (Urquhart and Weir, 1998).....	31
Table 2.	Numbers of the Students Participating in the Study.....	56
Table 3.	Majors of the Students Participating in the Study.....	56
Table 4.	Descriptive Statistics of the Paper-Based RCTE and the Computer-based RCTE in the Pilot Study.....	77
Table 5.	Correlation Matrix of the Paper-Based RCTE, the CB-RCTE, and the Entrance Examination Scores in the Pilot Study.....	79
Table 6.	Statistics Used in Selecting Items.....	80
Table 7.	Mean Scores and Reliability Coefficient of the Pilot Questionnaire.....	84
Table 8.	Design for Test Administrations.....	87
Table 9.	Descriptive Statistics of the RCTE.....	88
Table 10.	Correlation Matrix of the Paper-Based RCTE, the Computer- Based RCTE, and the Entrance Examination Scores.....	89
Table 11.	Descriptive Statistics and Reliability Coefficient of the Questionnaire.....	91
Table 12.	A Comparison of the Mean Scores from the Paper-Based RCTE and the Computer-Based RCTE.....	96
Table 13.	A Comparison of the Mean Scores of Part 2 of the Questionnaire Before and After the Test Administrations.....	98
Table 14.	The Interpretation of the Questionnaire.....	99

จุฬาลงกรณ์มหาวิทยาลัย

LIST OF FIGURES

	Page
Figure 1. The Program Outline of the Computer-Based RCTE.....	71
Figure 2. The Database of the Computer-Based RCTE.....	72



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER I

INTRODUCTION

Background of the study

The Office of the Higher Education Commission of Thailand (formerly known as the Ministry of University Affairs) has launched a National English Curriculum (Foundation English I and II) of tertiary educational level comprising two goals and seven standards as the guideline that every university offering English language courses must follow. The two goals were established to promote communicative skills among the students and the seven standards are criteria of what students should know and be able to do as a result of instruction (Outline of Standardized National English Curriculum at the Tertiary Level Report; unpublished). These standards can be set as main criteria to measure the students' English language abilities at the end of the instruction. In other words, teachers can establish standardized tests from these seven standards. Aiken (1998) stated that scores obtained from the achievement test cannot only be used in assigning grades but can also promote those students who did well to higher level courses. Less able students can also gain benefits from the achievement scores. The scores can identify their weaknesses and strengths which can lead to improvement and development in those areas. In other words, the achievement test can be used as a summative evaluation as well as an ongoing formative evaluation. Once every university in Thailand follows these national guidelines, it is possible that the country will need to administer a national achievement test of Foundation English. For high stake test situations, multiple forms

of the tests can and must be created. Together with conventional paper-and-pencil tests, computer-based forms could be promising.

In conjunction with the attempt to standardize the English curriculum throughout the nation, some universities have already begun establishing and offering online English courses similar to those offered in the main campuses to their provincial campuses around the country. In addition, electronic-learning (E-learning) has also provided opportunities for students to do extracurricular activities and supplements to regular classes at their convenience. The advanced features of computers and software allow teachers to create multimedia and interactive virtual classrooms that can enhance all four language skills and aspects of foreign language learning. Nowadays with the emerging cooperation between technology and teaching, the students are encouraged to learn autonomously and make use of electronic sources of information as much as possible while teachers become facilitators. This means that not only are students expected to have the necessary computer and English language skills to enhance their learning in the educational context, but also that external stakeholders such as parents and the business and industrial sectors expect them to possess operating and technical skills as well (Wongboonsin et al, 2003 cited in Prapphal, 2003). Panitchapakdi (1998 cited in Prapphal, 2003) affirmed that a workforce with technical capabilities will be highly needed. Those who are equipped with technical abilities will survive the strong competitiveness in the labor market. Because the learning mode is changing and becoming more demanding and the modern world is becoming more competitive, testing and assessing should be in tune with these changes.

The use of technology in the language classroom has been proved to be successful as a number of research studies have demonstrated. In the field of language assessment, the computer has also become a potential and effective tool in language testing. The use of computers in language testing started in 1935 when the IBM model 805 was developed to score objective tests and multiple-choice items (Fulcher, 2000). Chalhoub-Deville (2001) stated that language testers were familiarized with the use of technology about two decades ago with the emergence of the Language Testing Research Colloquium (LTRC) in 1985. At present, the availability of computers with advanced features and authoring software has been exploited to properly suit testing situations. From simple scoring machines, computers have now been adapted to administer advanced and high stakes tests such as adaptive tests, college admission tests, and online-placement tests. For example, the Test of English as a Foreign Language (TOEFL), a dominant proficiency test offered by the Educational Testing Service (ETS), the Computer-Adaptive Testing of Listening and Reading Comprehension of the Brigham Young University (Madsen, 1991), the Computerized Oral Proficiency Instrument (COPI) (Kenyon and Malabonga, 2001), the Web-based Language Assessment System (WebLAS) at UCLA (Bachman, 2003), and the IntelliMetric Automated Essay Scoring System (Edelblut, 2003) were studied during development to see how effective they are in assessing language abilities.

Technology, especially the computer, has assisted test developers in many ways. Burnstein, Frase, Ginther and Grant (1996) have provided a list of areas in which the computer and software have been used for language testing. The computer is used in test design, test construction, item trial, actual test item delivery, item

management, and score reporting. Brown (1997) also pointed out current and advanced uses of the computer in creating an item bank. By using commercially available computer programs, test developers can create, pilot, analyze, store, manage, and select test items that, later on, create multiple test forms. Hence, computers and the Internet are now powerful tools for testing and there is room for research in the use of this high technology in language assessment.

Statement of the problem

Researchers in reading have defined reading in many different ways. Urquhart and Weir (1998) admitted that they had difficulty in defining reading. However, after having reviewed definitions given by various researchers in reading, they defined reading as “the process of receiving and interpreting information encoded in language form via print.” Bernhardt (1999) viewed reading as “the extraction and construction of a message from a written text.” These two definitions suggest that reading involves cognitive processes which take place in the readers’ minds. Grabe (2000) defined reading in terms of a process accompanied by several essential features. These definitions of reading are based on studies and empirical research in the first language (L1). However, they can be used to explain reading in a second language (SL) or a foreign language (FL) (Grabe, 2000).

Reading in SL or FL is distinct from reading in the first language in that it involves constraints and other factors (Grabe, 2000). Bernhardt (1999) has suggested that in order to achieve understanding of the target language the readers need to be competent in their first language, have intention and motivation to read, and possess certain knowledge of that target language. Alderson (2000a) has also provided

factors or variables that can affect reading ability in SL and FL. The variables can be divided into two main categories—reader factors and text factors. Among the text variables, the medium of text presentation also plays a significant role in the reading process.

In language testing, there are additional factors that can influence the test-taking process and test performance. These factors, which are test method facets, personal attributes (of the test takers), communicative language abilities, and random factors, are unobservable but exist and can affect test performance as well as distort the true ability that the test is aimed to measure. Included in the test method facets are “familiarity of the place and equipment” and facets of the input (e.g. format of the input and mode of presentation) (Bachman, 1990:118).

Although the use of computers in language testing and the number of studies in this area is increasing, there are several concerns over this mode of test presentation. Fulcher (2000) pointed out important concerns about computer-based tests. Validity is the central concern in language testing (Fulcher, 2000; Green, 1988). Computerized tests are designed to measure the same constructs of certain abilities as conventional tests (i.e. paper-and-pencil tests). Hence, the two forms—computerized and conventional tests—should be equivalent. In addition to the aforementioned factors affecting test performance and the test scores, other variables such as test takers’ familiarity with computers, their affect and attitudes towards technology and other personal attributes can affect performance on computerized tests (Fulcher, 2000; Bachman, 2000).

Research studies that were conducted elsewhere to compare subjects’ performances on two modes of testing (i.e. computer and paper-and-pencil) reveal

two important findings. One is that, in some contexts, the subjects performed better on the computerized tests while in other situations they excelled in the paper-and-pencil tests. It seems that there are no conclusive results whether either mode of test presentation can facilitate test takers when taking tests. In addition to studying comparability of the two tests delivered via the two modes of testing, those researchers also investigated the subjects' preferences towards the two modes. The results from these studies revealed a positive tendency toward taking computerized tests even though, in some studies, the subjects reported they were more familiar with the paper-and-pencil counterpart.

In the Thai educational context, in which the English language is a foreign language taught widely in schools and institutions, language assessment, particularly in the use of computers in language testing has not yet been fully explored. Two studies were conducted to investigate interfaces such as the effects on reading of different font sizes and background colors on the screen (e.g. Gomontean, 1997 and Tetewornan, 1995). However, there is no study which compares the two modes of testing (paper-and-pencil and computer) in investigating English reading achievement of Thai students.

Purposes of the study

Therefore, this study aims to compare the test performances of Thai university students using a computer-based reading comprehension test of English and a conventional one. In addition, it compares students' attitudes towards the two modes of testing: the computerized and the conventional tests.

Research questions

In order to achieve the above purposes, two research questions were constructed:

1. To what extent can the computer-based reading comprehension test of English measure the reading skills of first year university students?
2. What are the students' attitudes towards the computer-based and conventional reading comprehension tests of English?

Statement of hypotheses

1. There is no significant difference between the mean score of the conventional reading comprehension test of English and that of the computer-based reading comprehension test. In other words, the mean score from the computer-based test of English is the same as that on the conventional test. ($\bar{X}_{\text{computer-based test score}} = \bar{X}_{\text{paper-pencil test score}}$)
2. The mean score of the attitudes towards the computer-based test is higher than that of the conventional test. ($\bar{X}_{\text{computer-based test attitude}} > \bar{X}_{\text{paper-pencil test attitude}}$)

Variables

The independent variable in this study is the two modes of testing (computer and conventional modes) while the dependent variables are as follows:

1. Scores from the conventional reading comprehension test of English
2. Scores from the computer-based reading comprehension test of English

3. Scale scores from the questionnaire on attitudes of the students towards the computer-based and the conventional reading comprehension tests of English

Scope of the study

In general, this study aimed to study the use of the computer in English language testing at the tertiary level. Concerning the feasibility and validity issues, certain aspects of the study were delimited. Regarding constraints from the subjects participating in the investigation, the focus group of this study was first year university students from the Faculty of Humanities at Chiang Mai University (CMU). Specifically, first year university students who were studying the compulsory Foundation English (FE) courses were selected. The participants' comprehension of passages in the English language was measured through conventional and computer-based tests. The tests were the Reading Comprehension Test of English (RCTE) and the other instrument was an attitude questionnaire. The researcher developed the two instruments.

Even though two different modes of test presentation were used, the tests were developed based on the same reading constructs and test specifications and were not adaptive tests. The reading passages were determined for their reading ability levels and examined for content appropriateness by three specialists. Following each passage were four-option multiple-choice test items. The tests delivered by the two modes of test presentation underwent analyses in order to establish their equivalence.

In the test development stage, the students' variables related to reading proficiency, such as schema and familiarity with the topics and technology were not

assessed despite the fact that they might affect the students' performance. Even though the participants' familiarity and experiences with the computer were not the main issues in this study, these factors were controlled and minimized in the computer-based test design. The tests were administered at CMU, where the subjects of the study were studying, in order to reduce the environmental and equipment effects on the test performance. The conventional test was administered in a classroom arranged as a testing situation and the computer-based test was held in a computer lab. The two rooms were located in the same building.

In short, this study investigated test performances of a group of first year university students from the Faculty of Humanities at CMU on English reading comprehension tests using two modes of testing and examined their attitudes towards the two modes. The instruments were developed by the researcher and the study was carried out at the Faculty of Humanities at CMU. Therefore, results of this study cannot be generalized to other situations.

Limitations of the study

The limitations of this study are in several areas: the test development, the sample selection, and the samples' familiarity with the computer.

The development of the Reading Comprehension Test of English (RCTE) is one of the limitations of this study and its use needs to be treated with caution. In the test development, the reading passages used in the test were selected based on the themes of Foundation English II (FE II) offered to first year university students at the Chulalongkorn University Language Institute (CULI). The reading skills measured were a combination of those specified in the test specifications of FE II and partly

from the models of reading in a second/foreign language. Therefore, it cannot be treated as a proficiency test but more or less as an achievement test. The aim of the test is to measure the reading abilities of the students in comprehending English texts of various topics. In addition, the test does not aim to measure how much the students who take FE II have learned in their class as a result of instruction. Because of this, it can be said that the RCTE was developed according to theory-based construct definitions (Bachman and Palmer, 1996: 117-120) specifically designed for this study. The test scores gained from these tests can only be interpreted as how well the students are capable of reading in the English language.

The sample selection also creates certain limitations for the study. The samples were not randomly selected. The population was limited to first year students from the Faculty of Humanities who were studying the Foundation English courses at CMU at the time of the study. They were purposefully selected from intact groups. In addition, not all students in those contact groups participated. They can thus be viewed as volunteer groups. In addition, they were not assigned to control and non-control groups as in the experimental method, so they can be viewed as a self-controlled group. Therefore, the results of the study might not be generalizable to a larger and more heterogeneous population.

Aside from the sample selection, certain characteristics of the sample may also contribute some effects on the results of the study. To categorize the English language proficiency levels of the participants, their English test scores from the Entrance Examination were collected. Because the researcher was unable to retrieve the students' actual official scores, self-reported scores, which might not be completely accurate, were collected. The categorization of the subjects into three

levels of English proficiency solely according to the English test scores from the Entrance Exam may not be appropriate in examining their reading proficiency levels since the scores did not report their reading ability in particular. In other words, the proficiency levels of the subjects reported in this study might not be concise.

The last limitation of the study is associated with the samples' familiarity with the computer. The subjects' familiarity with the computer is an important factor in doing the computer-based test. Technology anxiety can create cognitive workload (Noyes et al, 2004) and hinder the students' ability to do well on the test. As a result, the true test scores and true abilities being measured can be contaminated and inaccurate. The familiarity with the computer among the students participating in this study was not fully investigated prior to the RCTE administration. Therefore, the interpretations of the scores from the tests in this study should be treated with caution.

In short, even though the RCTE was written to measure reading comprehension skills, the schema, or the background knowledge, was not controlled in this study. The interpretation of the scores needs to be done with caution. Moreover, the selection of the samples to participate in the study creates certain limitations. The subjects of the study were purposefully selected from a limited population group. Thus, the results from this study might not be generalized to a larger and more heterogeneous population. The last limitation of this study concerns the subjects' familiarity with the computer. This characteristic of the subjects was not thoroughly investigated, although the subjects of the study revealed that they were able to operate the computer well in terms of basic programs such as using a word processor in typing reports and surfing the internet using web browsers (Internet

Explorer or Netscape). The possibility remains, though, that the subjects might not be familiar with computerized tests. As a result, the test scores could be distorted.

Definitions of terms

The following are operational definitions of terms that are used in this research report.

English reading achievement

English reading achievement was measured from the Reading Comprehension Test of English (RCTE). The Reading Comprehension Test of English is constructed to measure reading comprehension skills in the English language. It is based on constructs and theories of reading in a second or foreign language proposed by reading experts such as Munby (1978) and Grabe (2000). The reading comprehension test is composed of four reading passages on topics related to technology and alternative therapies. The passages are examined for readability levels using the Flesch Readability formula available in Microsoft Word Processor. Each passage is followed by multiple-choice test items. In addition to the reading comprehension part, there are vocabulary test items to measure the students' ability to guess word meaning from the given context. Each test item is composed of four options. Each correct answer is worth one point. The scores from each mode are accumulated and analyzed separately and the results are used to test the hypothesis. (See Chapter III for the complete development of the RCTE.)

Modes of testing

“Modes of testing” in this study refers to the methods that are used to deliver and administer the reading comprehension tests of English. The modes of testing are the conventional (paper-and-pencil) and the computer modes.

Conventional test

“The conventional test” refers to the test that is printed on paper. The RCTE, for example, is in booklet form. In the term “paper-and-pencil test,” “paper-pencil,” and “paper-based” are used interchangeably in this report. The subjects answer test items by choosing their desired option for that item on an answer sheet provided.

Computer-based test

“The computer-based test” refers to a version of the RCTE that is installed in a computer program specially designed for this study. The test is delivered in the same linear fashion as the conventional test. In other words, the computer-based test is not an adaptive type. The test appears on a computer screen. The participants respond to the test by using a mouse. In the reading comprehension section, a reading passage appears on one side of the screen and a test item appears on the other side. There is a scrolling bar to allow the students to scroll up and down while reading the passage. The reading passage remains till the last test item for that passage is selected. One test item appears at a time. The students click an option they decide on and navigate to the previous or next test items using the navigator icons provided. The students are allowed to move back and forth and change their answers in the same way as in the conventional test. Once the students finish, the program calculates the score and presents the test score report.

Attitudes towards testing using the computer and conventional modes

“The attitudes” in this study refers to favorable or unfavorable opinions on the two modes of test presentation. The attitudes of the students towards the test using the conventional and the computer modes are measured from the questionnaire using a five-point Likert’s scale. The scale of one represents the least positive attitude while the scale of five represents the most positive one. The scores are accumulated and analyzed statistically for, for example, frequency, mean scores, and correlations.

Significance of the study

The results from this study are beneficial to education in many ways. In terms of test development, a computer-based reading comprehension test of English has not been constructed at the university level in Thailand. Together with the results concerning the students’ attitudes towards the computer-delivered test, the results of the study can provide teachers with some considerations about the implementation of computer-based tests. Besides, teachers or lecturers especially those in large institutions and those who are involved in developing in-house tests and researchers should be aware of the standardization processes in test development and test administration, especially the variation of forms of a test and modes of test presentation.

If this study reveals the results as hypothesized, it can be inferred that the two tests are equivalent and computers can be used to deliver tests with minimal effects on test performance. The methods of test presentation, especially the computer, may not have any impact on test performance. Then, the computer-based version can be further developed and used interchangeably with the conventional one. By doing this,

test developers and students can gain quite a number of benefits. For the test developers, new test items can be added in the program and tried out with the existing ones. This means they do not have to spend time and resources developing an entire new set of tests. In addition, the test writer can develop an item pool. Also, it takes less time to administer and do the computer-based test. The students can arrange the test time and date to suit their convenience. This can reduce the cost of administering a test and the budget for printing out test booklets.

Moreover, computer-based tests can be generated and further developed to be more advanced and efficient adaptive tests that can be administered as high stakes tests to a large number of examinees. In addition, the results of this study may provide some details or answers concerning controversies about the effects of the use of technology in language testing particularly in the context of English as a second or foreign language.

Apart from providing benefits in the educational setting, stakeholders in the private sector, companies, and employers can also gain benefits from this study. In the present competitive business world computer skills and foreign language competence are required. Stakeholders can use computers to measure language abilities as well as necessary computer skills in selecting or recruiting new staff.

However, there is a possibility that the findings may not be as hypothesized. If the test result on the paper-and-pencil test outdoes that on the computer-based one and if the students show negative attitudes towards the use of computers in language testing, the results of this study might yield some insights about the use of technology in language testing.

In short, the results of this study will assist educators in introducing computers to the assessment processes and language instruction. The business and industrial sectors can also employ the results of this study to develop tools for measuring and evaluating their workforce. Moreover, the findings of this study can give information about the use of computers in language testing particularly in testing reading comprehension in Thailand.

Summary

Chapter I begins with a brief background of the national Foundation English (FE) courses at the university level and an overview of the use of technology, especially the computer, in language testing. The issues to be investigated in this study are then stated. Included in the chapter are the purposes, research questions, hypotheses, variables related in the study, scope and limitations of the study, definitions of terms, and the significance of the study.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter reviews studies on models of reading and reading comprehension, applications of computers in language testing, and modes of testing. The chapter ends with a review of studies and research related to comparing conventional and computerized tests.

Reading models

Models of reading are an attempt to explain what is happening while a person is reading. In other words, models of reading concentrate on the process of reading. These models aim to clarify how a reader comprehends a text being read. Elements or components of these processes are reviewed and used to establish grounds for assessing reading, particularly in a foreign language. The models of reading can be categorized into three major groups: bottom-up or text-driven models, top-down or reader-driven models, and interactive models. The three models are described below.

1. Bottom-up models

The major figures in developing these models are Gough (1972), LaBerge and Samuels (1974), and Carver (1977) (Barnett, 1989). These models describe the reading process as sequential activities progressing from lower-level to higher-level cognitive stages. The lower-level stages are in the order of linguistic development starting from the smallest units: letters, words, phonological systems, phonemic representations, syntactic rules and semantic systems. These stages are activated to comprehend a sentence (Rayner and Pollatsek, 1989). The process involves decoding

and the reader is seen as a “passive decoder” (Barnett, 1989; Alderson, 2000a). However, these decoding processes take place automatically and quickly (LaBerge and Samuels, 1974 cited in Barnett, 1989; Rayner and Pollatsek, 1989).

These models give priority to the text and its characteristics being read and the reader is to decode meaning from the text. The reader’s skills and linguistic knowledge, namely graphemes, phonemic knowledge, syntactic knowledge, and semantic knowledge contribute to comprehension. Hence, the bottom-up models emphasize comprehension at the word and sentence level (Wallace, 2001).

2. Top-down models

These models of reading emphasize the reader’s variables which play significant roles in the reading process and comprehension. The proponents of these models are Goodman (1968) and Smith (1982) (Barnett, 1989). Goodman (1967 cited in Barnett, 1989; Rayner and Pollatsek, 1989) explained the reading process as a “psycholinguistic guessing game” in which the reader possesses certain world knowledge and contextual knowledge which allow the reader to form some hypotheses about the text to be read. The reader then relates the knowledge or hypotheses to the information extracted from the materials being read. In other words, the reader scans the text and selects information using his or her prior knowledge of the language, strategies, and cognitive styles. He or she then forms an integrated “image” from the knowledge he/she brought to and the information from the text, partly to confirm and partly to make further predictions (Goodman, 1967 cited in Urquhart and Weir, 1998). Moreover, Goodman (1968 cited in Barnett, 1989) stated that while reading the reader makes use of his/her knowledge of syntax and semantics rather than depending on the print and phonic system of the text.

Rayner and Pollatsek (1989) observed that these processes happen in a cyclical fashion. Barnett (1989) summarized that, under Goodman's view on reading, the processes involved while reading are predicting, sampling, confirming and correcting.

Alderson (2000a: 17) suggested that these models of reading could also be called "schema-theoretic models" since the models emphasize the roles of the knowledge that the reader brings to comprehend the text. He also noted that "they are based upon schema theory, which accounts for the acquisition of knowledge and the interpretation of text through the activation of schemata: networks of information stored in the brain which act as filters for incoming information."

Urquhart and Weir (1998) suggested the term "reader-driven" to describe these reading processes since the reader applies his/her knowledge together with information from the text in order to comprehend the message. The readers can thus be viewed as active readers since they use their knowledge to interact with the information encoded in the text (Wallace, 2001).

3. Interactive models

Grabe (1988: 59) stated that "interactive models of reading assume that skills at all levels are interactively available to process and interpret the text." The interactive process of reading should not be viewed as interaction between the reader and the text being read. Barnett (1989) briefly summarized that the interactive models view the reading process as one in which "high-level processing stages influence lower-level stages" and elements of the texts being read are major components in comprehension.

The major figures in developing these models are Rumelhart (1977), Stanovich (1980), Rumelhart and McClelland (1986), and Just and Carpenter (1980).

Rumelhart (1977 cited in Barnett, 1989) hypothesized that the reader makes use of his/her knowledge of syntactic, semantic, lexical, and orthographic knowledge rather than relying mainly on perception of the print. The reader stores words that have similar meanings or related uses and then links them into other larger propositions. The process of reading starts from seeing words which are then stored or memorized. This process is called “visual information store”. What happens in this process is that features of the words are extracted and these patterns are synthesized by using prior knowledge (i.e. knowledge about spelling, patterns, syntax, vocabulary, semantics, and context). From this integration, the reader interprets the text and achieves understanding. However, the stages of the process can be random and varied depending on the context of the text and the availability of information sources. Rumelhart thus summarized that “all of the various sources of knowledge, both sensory and nonsensory, come together at one place and the reading process is the product of the simultaneous joint application of all knowledge sources” (Barnett 1989: 23).

Stanovich (1980 cited in Barnett, 1989: 26-27) proposed a model as an extension of Rumelhart’s model called the “interactive-compensatory model”. Referring back to the bottom-up and the top-down models, in the bottom-up models, the readers rely mainly on the lower level processes. According to the top-down models, the readers rely on the higher level processes. With the interactive-compensatory model the readers might make use of other strategies or other “knowledge sources” to overcome the difficulties in reading regardless of the order of the levels of the processes. For example, less able readers may use syntactic or semantic knowledge to compensate for the lack of lexicon knowledge while

competent readers may apply reading skills in first language instead of linguistic knowledge of the target language. Similar to Stanovich's model, Rumelhart and McClelland et al. (1986 cited in Barnett, 1989) offered their model called Parallel Distributed Processing (PDP). This model presupposed that every source exists in the mind and they all interact with each other in such a way that the more effective processes contribute information to one another in the process.

Grabe (1988: 56-70) claimed that the interactive models of reading have apparently been well accepted in the context of English as a second or foreign language (ESL/EFL) because they give importance to both higher-level and lower-level processing abilities in the reading process. In addition, high and low ability EFL/ESL readers are believed to be able to compensate for reading skills that they are not good at in order to overcome difficulties while reading. However, ESL/EFL readers need to possess a wide range of vocabulary and accessibility to those words.

The aforementioned models of reading explain the development of research and models about processes of reading in general without putting emphasis on how the reader of one language reads in another language. In addition, knowledge of target language also plays a crucial role. The knowledge of the language is linguistic knowledge of that language starting from low-level knowledge to high-level knowledge. The bottom-up and the top-down models appear to explain ways or style readers approach a text. Some readers may approach the text by concentrating on individual words to form meaning. Some readers may concentrate on understanding the meaning of the text and pay less attention to minor components of the target language. The interactive models describe the process of reading and give importance to all knowledge of the target language. In the process according to the interactive

models, every source of knowledge of the language interacts with every other. It doesn't mean that the readers are very competent in that target language. It means every source of knowledge contributes its knowledge to the process to help the readers reach understanding or meaning. From all three major models, it can be summarized that the way the readers approach the text depends on the strengths of their knowledge of the target language. The following section presents similarities and differences between reading in the first language and reading in the second or foreign language.

Reading in the first (L1) and the second/foreign language (SL/FL)

1. Processes of reading in L1 and the SL/FL

Barnett (1989) explained that the process of reading in a second/foreign language (SL/FL) could be categorized as the same process described in the interactive models for first language reading. In addition, knowledge about the language and knowledge about the world that the reader possesses interact with the information from the text being read in order to arrive at meaning.

Reading in SL/FL concerns the reader's variables that contribute to the reading process and the result of reading is comprehension. Reading in SL/FL can be regarded as a reader-based process. The variables or factors of the reader that interplay in the reading process are "purposes, cognitive skills, language proficiency, strategies, background knowledge and schemata" rather than the variables from the text such as "graphic, syntactic, and semantic features" (Barnett, 1989: 36; Bernhardt, 1984; Carrell, 1998: 103). To explain the reading process in SL/FL, two existing

models have been proposed, namely Coady's psycholinguistic model (1979) and Bernhardt's (1991) constructivist model.

1.1 Coady's psycholinguistic model

Coady (cited in Barnett, 1989: 39-40) believed that the result of the reading process is comprehension, which is achieved "from the interaction of conceptual abilities, background knowledge, and process strategies." His proposed process strategies included:

- Phoneme-grapheme correspondences
- Grapheme-morpheme correspondences
- Syllable-morpheme information
- Syntactic information
- Lexical meaning and contextual meaning
- Cognitive strategies
- Affective mobilizers

In this model, less able SL readers make use of "concrete process strategies" while more able SL readers rely on more complex process strategies such as syntactic and semantic ones and sample the text to affirm the prediction that they create about the text being read. The reader may use different strategies when approaching different kinds of text and s/he can use several strategies simultaneously and switch to other strategies when the ones being used cannot lead her/him to comprehension.

1.2 Bernhardt's constructivist model (1991)

The reading process in this model is said to be an interaction between "text-based" and "extratext-based" components. The text-based elements are "word-recognition, phonemic/graphemic decoding, and syntactic feature recognition. The

extratext-based components are intratextual perceptions, prior knowledge, and metacognition” (Barnett, 1989: 47-48). Those elements actively interact with one another while the reader reads a particular text.

As mentioned earlier, SL/FL readers are believed to activate not only knowledge of the language and its linguistic properties within the texts they read but also their world knowledge or schema to help them achieve comprehension.

1.3 Second/Foreign language reading and schema theory

Schemata are simply defined as “a reader’s existing concepts about the world” (Anderson and Pearson, 1984: 255 cited in Barnett, 1989: 42). Rumelhart (1977 cited in Barnett, 1989: 43) explained that the extent to which the schemata will be used or activated depends on the amount of information in the text. In addition, the reader may use more than one schema to approach the text. Each individual reader may use schemata differently. There are two types of schemata: content schemata, or background knowledge, and formal schemata, or knowledge about text structure or rhetorical organization. In addition, schemata are related to culture as well (Barnett, 1989: 43).

Research cited in Barnett (1989: 53-56) agreed that processes of reading in the first and in second language are very similar. To comprehend the text in the second language the readers should be competent in that language rather than rely on their first language reading abilities. A lack of adequate knowledge of the second language may impede the reader from creating meaning from the text, and to circumvent the problems the reader tends to rely on low-level components in reading. The first language structures and grammar may also interfere with the process of reading in the second language. However, the first language reading skills are believed to transfer to

reading in the second language. Experienced readers in any language tend to employ similar reading processes when reading other languages, while less able readers use different processes when reading. In addition, cognitive strategies and logical reasoning skills seem to enhance reading comprehension; in contrast, relying on a graphic representation of the text may not help readers understand the text.

2. Distinctions between reading in the first and the second/foreign language

Although there have been proponents of the view that processes of reading in the second/foreign language (SL/FL) are similar to those in the first language (L1), there are still certain significant factors that make SL/FL reading distinct from reading in the native language. Those components are important because they can enhance or hinder the process of reading a new language.

Grabe (2000) has illustrated “constraints” that make SL/FL reading different from L1 reading. The major concerns about reading in SL/FL are related to knowledge of the target language. The SL/FL reader needs to acquire a wide range of vocabulary in that target language. Bernhardt (1991: 5) also pointed out that there are differences in linguistic bases between the first and the second/foreign language. Those linguistic features are syntactic representation and phonetic, semantic, and rhetorical information. Grabe (2000) stated further that the closer the relationship between orthographic representations and phonological properties, the more able SL/FL readers might be. There is also a language threshold, which can be explained as difficulties accompanying learning to read another language that SL/FL readers need to overcome in order to be able to read fluently. Besides the knowledge of the target language, cultural knowledge of that language is also crucial in helping the

SL/FL readers achieve understanding from the material being read. The extent to which the readers are exposed to the target language materials and uses also contributes to the level of fluency that the SL/FL readers will obtain. According to Grabe (2000), some components in the reading processes such as word recognition, semantic information, and world knowledge appear to transfer once the readers acquire reading skills in their native language. However, due to the limited range of vocabulary and knowledge of the target language, the SL/FL readers may not be able to read quickly. This may delay the whole process in reading. One last important distinction is the motivation of the reader in reading the second/foreign language.

Similar to Grabe, Alderson (2000a: 23) added that “knowledge of the second language (or the target language) is a more important factor than first-language reading abilities”. Once reading abilities in the native language are acquired, learners can learn to read in the second/foreign language. What the second/foreign language readers need to develop and improve is knowledge of that second/foreign language (Cummins, 1979, 1991, cited in Alderson, 2000a: 23).

In short, the reading process in a SL/FL is similar to the process of the interactive models in the first language. All sources of knowledge about the target language contribute to the process. Together with the linguistic knowledge of the target language, schema, or background knowledge, and knowledge of the culture of the target language are also important to enable the readers to understand the text well.

Reading comprehension

Comprehension can be defined as “the act or ability to understand spoken or written text” (Davies et al, 1999). Urquhart and Weir (1998) defined comprehension as the process of getting information from written texts. Similarly, Rayner and Pollatsek (1989) referred to comprehension as the ability “to extract meaning of the text that is being read.” There are also influences of background knowledge on comprehension. Different readers may achieve different comprehension. In other words, the readers may comprehend the same text differently.

However, comprehension is not viewed as being the same as interpretation. Urquhart (1987 cited in Urquhart and Weir, 1998: 112-115) explained differences between comprehension and interpretation. According to Urquhart, readers can achieve comprehension differently due to their intentions and purposes. The results of this are different levels of comprehension. On the other hand, interpretations are influenced by different information, such as background knowledge and proficiency in the target language. They can interpret the text differently and create different results. Readers may interpret the same text differently when reading it at different times because they may carry different thoughts. Therefore, comprehension varies depending on the extent to which factors such as background knowledge and purposes come to play in the reading process.

Readers cannot only comprehend the text but they also have to understand different interpretations which are influenced by different schemata. Such differences and effects should be minimized, especially in assessing reading comprehension. Urquhart and Weir (1998: 112-114) suggested three ways to limit the roles of schema in designing reading tests. One way to minimize the effects of schema is to provide

short texts from various fields or topics regardless of the examinees' backgrounds. This means some examinees may be familiar with the selected texts while some are not. Another way is to select texts that are not related to the examinees' fields of study or backgrounds at all. The last solution is to survey for compromising areas of interest among the examinees. By doing this, most examinees receive equal opportunity to access the selected texts.

Grabe and Stoller (2002: 14) suggested that comprehension is one of the purposes that encourage people to read. Comprehending texts is the common and basic purpose in reading and it is more complicated than other kinds of reading. They also stated that "reading for general comprehension requires very rapid and automatic processing of words, strong skills in forming a general meaning, presentation of main ideas, and efficient coordination of many processes under very limited time constraints" (p.14). Their process is composed of ten processes involved in (fluent) reading comprehension, namely a rapid, an efficient, an interactive, a strategic, a flexible, an evaluating, a purposeful, a comprehending, learning, and a linguistic process (p. 17). In addition to the linguistic process, the readers should also be competent in the language. Bachman (1990: 87) suggested that components of language competence are grammatical, textual, illocutionary, and sociolinguistic competence. Grabe's (1991 cited in Alderson, 2000a: 13) components in fluent reading are based on knowledge of the language: vocabulary and structural knowledge, formal discourse structure knowledge, content and world background knowledge, and metacognitive knowledge (i.e. recognizing important information, adjusting reading rate, skimming, previewing, using context to clarify understanding, and formulating questions about information). Reading comprehension is, hence,

viewed as an interactive process where the various processes listed are executed at the same time. Each individual reader may reach different levels of comprehension.

Although education practitioners and teachers believe that reading comprehension is the ultimate goal that the learners are expected to achieve, comprehension is only a part of the reading processes. Comprehension occurs in a continuum. Alderson (2000a: 7-8) pointed out that there are three levels of understanding. The first one is literal understanding. Another is “an understanding of meanings that are not directly stated in text” which can be termed “referred meaning.” The last level is “an understanding of the main implications of the text” in which the readers employ critical thinking about the text being read.

In sum, reading comprehension can be described as an act to extract meaning from a text and integrate that meaning with various sources of knowledge such as cultural knowledge, schema, and linguistic knowledge of the language of the text. Individuals may comprehend the same text differently because of the influences of the various sources of knowledge mentioned above. Therefore, it can be said that reading comprehension occurs in a continuum.

Reading skills in the second/foreign language (SL/FL)

From the aforementioned definitions and processes of reading comprehension, it can be said that to achieve comprehension in reading in the SL/FL context the readers need to have knowledge of the target language in terms of linguistic knowledge, organization, and socio-cultural aspects of that language. Together with knowledge, the readers also need to employ skills in reading.

In the study of reading there are two terms (i.e. reading skill and strategy) which are sometimes used interchangeably since there is some overlap between the two. The definition and distinction between a reading skill and a reading strategy is not very clear (Grabe and Stoller, 2002). Urquhart and Weir (1998: 96-98) believed that “strategies are reader-oriented” while “skills are text-oriented.” They explained further that the skills are more automatic while the strategies are used under the readers’ control. Strategies are executed to circumvent problems which in this case are those related to reading difficulties such as guessing unfamiliar word meanings. They also defined a reading skill as “a cognitive ability which a person is able to use when interacting with written texts” (p. 88) and they mentioned that skills are a part of the reading process.

There has been an argument whether or not reading is composed of and can be assessed on separate skills. Concerning reading comprehension, there are two different proponents. One group proposed the Unitary Hypothesis, which explains reading as a single act. The other group (i.e. the Multidivisible Hypothesis) believes that reading is composed of separate skills (Urquhart and Weir, 1998: 92). Although reading is composed of low-level and high-level components involving the interaction between skills and knowledge, it should be viewed as a combination of several sub-skills. To help learners decode the text, teachers should help them develop those skills and strategies. In terms of assessment, such skills and strategies should be considered when developing reading tests.

Urquhart and Weir (1998: 122-123) provided a broad organization of reading skills under four different levels. They called it a “matrix of reading types” which can be illustrated as follows:

Table 1. Matrix of Reading Types (Urquhart and Weir, 1998)

	Global	Local
Expeditious	A. Skimming, searching information, understanding signified information	B. Scanning
Careful	C. “Reading carefully to establish accurate comprehension of the explicitly stated main ideas the author wishes to convey; prepositional inferencing.”	D. “Understanding syntactic structure of sentences and clauses. Understanding lexical and /or grammatical cohesion. Understanding lexis/deducing meaning of lexical items from morphology and context.”

Davis (1968 cited in Alderson, 2000a: 9-10) provided eight broad skills involved in reading:

- “recalling word meanings
- drawing inferences about the meaning of a word in context
- finding answers to questions and answering explicitly or in paraphrase
- weaving together ideas in the content
- drawing inferences from the content
- recognizing a writer’s purpose, attitude, tone and mood
- identifying a writer’s techniques
- following the structure of a passage”

Munby (1978, cited in Alderson 2000a: 10-11) provided these micro skills in reading.

- “recognizing the script of a language
- deducing the meaning and using of unfamiliar lexical items

- understanding explicitly stated information
- understanding information when not explicitly stated
- understanding conceptual meaning
- understanding the communicative value of sentences
- understanding relations within the sentence
- understanding relations between parts of text through lexical and grammatical cohesion devices
- interpreting text by going outside it
- recognizing indicators in discourse
- identifying the main point or important information in discourse
- distinguishing the main idea from supporting details
- extracting salient details to summarize the text
- extracting relevant points from the text selectively
- using basic reference skills
- skimming
- scanning to locate specifically required information
- transcoding information to diagrammatic display”

Bloom and colleagues established a taxonomy of the cognitive domain which is composed of a hierarchy of abilities (Bloom et al. 1974). The taxonomy is organized from simple to complicated “behavior” and it comprises six major categories: knowledge, comprehension, application, analysis, synthesis, and evaluation. Bloom believed that those ability domains are acquired sequentially. The abilities are then accumulated and integrated.

From the previous section, it can be said that to be able to read fluently in the second/foreign language the readers need to be competent in that language as well as be able to apply reading skills to approach the texts. Several types of knowledge and reading skills interact with each other to help the readers achieve comprehension. In order to prove whether the SL/FL readers receive the intended message from the text being read and to facilitate teachers and learners in improving their reading ability, reading skills that are essential in SL/FL reading should be provided. SL/FL readers often encounter problems while reading or they may not fully achieve the intended understanding of the texts. There are many factors that can affect the reading process and hinder the reader from comprehension. Important factors contributing to the reading process and comprehension are reviewed in the following section.

Factors affecting reading in the second or foreign language (SL/FL)

Alderson (2000a: 32-84) compiled factors that contribute to the processes of reading in the second or foreign language (SL/FL) and the factors are divided into two main groups: the factors from the reader and the factors from the reading materials.

1. Factors related to the reader

1.1 Roles of schemata

Since reading can be viewed as a cognitive process, thoughts and knowledge that the reader possesses will certainly play a role in the reading process. The existing knowledge that the reader has is called schemata. The information retrieved from the new text that is being read will be integrated with the existing schemata if they are matched or related; otherwise, the new information will be ignored. The schemata also have a role in selecting what new information will be recognized and stored.

There are two types of schemata: formal schema and content schema. The formal schemata refer to knowledge about “the language and linguistic conventions,” including knowledge of text organization and types of genre (Carrell, 1983a cited in Alderson, 2000a: 34). Studies in SL/FL reading focus on syntactic, lexical, rhetorical, and metalinguistic knowledge (awareness of the linguistic knowledge). In first language reading research, the larger the reader’s vocabulary size, the more comprehension he/she will achieve. In SL/FL reading research, the syntactic and vocabulary knowledge of the language will enable the reader to understand detailed information. Besides, knowledge of how the text is organized and uses of signal words and information will help the SL/FL readers understand the text well. If the readers know how to monitor their knowledge of the language and monitor their uses of reading strategies to overcome reading difficulties, they will be able to understand the text better. Alderson (2000a: 39) then concluded that “second language knowledge is more important than first-language reading ability.”

Another type of schemata is content schemata, which can be divided into two subcategories. One is background knowledge which “may or may not be relevant to the content of a particular text.” The other is “subject-matter knowledge which is relevant to text content and topic” (Alderson, 2000a: 34). The reader needs to have knowledge about the text in order to understand it well. If the reader becomes familiar with texts sharing similar themes, s/he can understand those texts easily. There are research studies showing that cultural knowledge of the language of the text being read also contributes some effects to the reading process and understanding.

1.2 Reader motivation and affect

Alderson (2000a) pointed out that poor readers in the SL/FL context often lack interest or motivation to read. There are two types of motivation: intrinsic and extrinsic. The extrinsic motivation urges the reader to read “at a surface level paying attention to facts and detail” (p. 53). On the other hand, with intrinsic motivation the reader will read carefully, concentrating on important and detailed information and making inferences as well as evaluating the text. In other words, when the reader is intrinsically motivated his/her high-level processes will be activated.

Reader anxiety also plays a crucial role in test performance and this can affect actual test scores. Besides, anxiety is connected to the reader’s motivation. Fransson (1984 cited in Alderson, 2000a: 54-55) found that learners who appeared to have anxiety during the tasks revealed that they also had low intrinsic motivation. Those who suffered from habitual anxiety or had temporary anxiety while doing the tasks appeared to do less well in the test.

1.3 Beginning readers and fluent readers

There are two significant points regarding less fluent and fluent readers: (1) different uses of skills among good and poor readers and (2) essential/necessary skills and knowledge to help readers become fluent ones. In SL/FL reading the reader needs to develop not only reading ability but also language proficiency (Alderson, 2000a: 60). He/she also needs to be able to use metacognitive skills effectively. The important metacognitive skills are:

- “recognizing the more important information in text
- adjusting reading rate
- skimming

- previewing
- using context to resolve a misunderstanding
- formulating questions about information
- monitoring cognition, including recognizing problems with information presented in text or an inability to understand text”

2. Factors related to the text

2.1 Topic and content

Again, the topic and content of the text, especially in a reading test, tends to facilitate those readers who are familiar with the topic and content areas while those who are not acquainted with the content may do less well. This shows that there is a close relationship between schemata and the topic and content of the text. Bachman and Palmer (1996: 120-127) suggest that in designing tests that are good and fair to all examinees, three approaches should be considered: (1) background knowledge related to that particular text should be excluded and not be assessed at all, (2) background knowledge can be included and assessed alongside language ability, and (3) background knowledge should be treated separately from the language ability to be measured.

2.2 Text type, genre, and organization

Certain text types and genres also facilitate or slow down the process of understanding. Expository texts tend to be more difficult to read than narrative texts. A text that is well organized in terms of the connections between ideas, logical sequencing, and rhetorical structures will help the reader achieve comprehension better than a poorly organized text.

2.3 Traditional linguistic factors

Linguistic factors, namely syntactic structures, discourse, vocabulary, topic and content, all contribute some effects to the reading processes. Differences in syntactic and discourse structures across languages may obstruct the readers. They turn to low-level processes in reading and recognize function words rather than content words (Alderson, 2000a: 69). As a result, reading comprehension will be delayed.

2.4 Text readability

Readability is explained as “the degree to which a given group of people find certain reading matter comprehensible” (Davies et al, 1999: 163). Text readability can be calculated using readability formulas in order to select appropriate texts for a reading comprehension test. Text readability also deals with the crucial issue of text authenticity. Any texts used in a test situation may be simplified by changing syntactic structures and lexis. However, studies have shown that simplification of texts may distort test performances and actual test scores (Alderson, 2000a).

2.5 Other related factors

Other related factors that may enhance or impede reading comprehension are typographical features and verbal and non-verbal information. Font sizes and clarity as well as non-verbal information such as graphs or illustrations may affect the reading process in some way. Furthermore, the medium of test presentation, especially on a computer screen, as suggested from various studies, also has some impact on the reading processes and performance (Alderson, 2000a: 78-79).

In sum, reading comprehension can be viewed as a personal act that is enabled by many sources of knowledge which are believed to exist in the mind of the readers.

It can be said that many knowledge sources intermingle in the process of reading. This, in turn, allows reading comprehension to be measured either as a unitary act of combined sources of knowledge or as composed of discrete reading skills. Despite the two opposing views about reading comprehension, the notion that reading comprehension is composed of individual reading skills enables assessors to measure reading comprehension.

Besides the integration of processes and knowledge, many factors both from within the readers and from external sources also have an impact on the comprehension process. Such factors should be considered in the test development process. Among those factors, the effects of the medium of text presentation are also significant to the reading comprehension process. The following two sections are reviews of the use of the computer in assessment and of studies on the effects of the use of computers in the assessment of reading comprehension.

Applications of computers in language testing

The use of the computer in the context of language teaching and learning is not new. The computer has been used to aid teachers for quite a long time. At present the use of the computer in this context is very fruitful and apparently effective. E-learning (electronic learning) and distance learning have reached a vast number of students and learners regardless of age, race, and location. It seems that learning and teaching modes have been changed. Students or learners may spend more time reading articles or completing assignments on the computer while the time in the classroom is used for discussion and face-to-face interactions if needed.

However, the use of the computer may be new in language testing. The introduction of the computer-based Test of English as a Foreign Language (TOEFL) by the Educational Testing Service (ETS) in 1998 first brought the issue to the attention of scholars and experts in the field and even of test takers. Since then the computer and its advanced technology have been exploited and applied in language testing. It appears that people involved in language testing can gain a number of benefits from this high technology. However, there are also some drawbacks that should not be neglected.

1. Advantages of the use of the computer in language testing

The advantages of using the computer and its technology in language testing and computerizing tests are mainly related to test administration. Two groups of people involved in testing, test administrators and test takers, can gain benefits from computerized testing.

1.1 Benefits for test developers

Brown (1997) pointed out that using the computer in testing allows test administrators to score the test effectively and to report the results accurately. In addition, some programs allow test administrators to provide the test takers with immediate feedback in terms of reporting scores and diagnostic feedback. Alderson (2000b) clarified that immediate feedback seems to enhance learning since test takers or learners are informed about their strengths and weaknesses. With the availability of the item response theory (IRT) statistics, testers are able to measure true abilities of test takers (Brown, 1997). The testers can create many items and store them in an item bank or item pool. The item banking, then, enables the testers to create multiple test forms that can limit practice effects and cheating (Brown, 1997). The tests can be

stored and delivered in different ways: CD-ROM-based and diskette-based or Internet/Web-based (Alderson, 2000b). The CD-ROM-based testing is flexible in terms of test scheduling and individualized administration. It also allows immediate feedback. However, test items are fixed and are not easily updated. Internet or Web-based testing allows the testers to put a large number of test items into the database which can be updated frequently. Through this method the test developers can also try out new test items together with other existing items. Scores can be sent directly to score users right after the test is completed. However, the security of this kind of test delivery is uncertain.

Chalhoub-Deville (2001) projected some benefits related to test validation processes. In her opinion, technology also enables test developers/testers to create interrelations among task characteristics, test takers' performances, and inferences about intended underlying abilities and processes which will give more precise details about test takers' abilities. New authentic task types that can draw out underlying abilities can also be generated. In other words, technology can be used to draw out true underlying abilities through its advanced ability to create authentic task types.

1.2 Benefits for test takers

Brown (1997) pointed out that students or the test takers can work at their own pace and they may not spend much time completing a test. They may be less frustrated since in some test programs test items are selected according to the competence levels of the test takers. They may also be less overwhelmed since one single test item may be presented at a time. But some may be more frustrated because they cannot predict how difficult or easy the next item will be as in the case of adaptive tests. Alderson (2000b) added that computer-delivery tests can be viewed as

user-friendly since they can provide a wide range of support such as visual or verbal clues.

2. Disadvantages of the use of the computer in language testing

For computer-based testing, at the moment, there are limited types of test items. Test developers who intend to deliver their tests through the computer often make use of an easy-scoring test item type: multiple choice (Alderson, 2000b). Brown (1997) categorized the disadvantages into physical and performance aspects. His physical disadvantages of computerized testing deal with the availability of computers for administering the test. Computer screen capacity for long reading passages and its graphic capacities also weigh against the use of computers in testing. The disadvantages related to test performances concern the effects of different modes of test presentation or the effects of the degree or level of computer familiarity among test takers. Alderson (2000b: 594) claimed that “reading from a screen is not the same thing as reading from print, and the need to move to and fro through a screen is much more limiting than being able to flick back and forth through print.” The test takers may have difficulty working on the computer and some may have anxiety about using technology. In some computerized test programs, providing immediate feedback after each item is answered can lead the test takers to change their answers. In doing this, true ability that is to be measured may be distorted (Alderson, 2000b).

3. Considerations in the application of the computer in language testing

Despite the increasing use of computers to deliver tests, issues about this mode of test presentation are essential and have been raised among the testers. The issues are similar to other issues in traditional test development processes. These issues

concern the use of the computer in testing language abilities in general. However, some issues are directly related to computer adaptive testing rather than less complicated computer-based testing.

3.1 Issues concerning test purposes and constructs

The second language construct is multidimensional and involves a variety of interacting components and processes (Bachman, 1990; Bachman and Palmer, 1996 cited in Chalhoub-Deville, 2001). Fulcher (2000) also pointed out that the use of computers in language testing is not yet fully developed or acceptable due to certain reasons related to test development procedures. Specifically, there is concern about how language testers measure integrative language abilities through the use of computers, since at the moment receptive skills are being measured but not productive skills (Brown, 1997). Moreover, Fulcher (2000) raised the ethical aspect of computer-based testing. Since a computer-based test measures the same traits as does a paper-pencil test, the two modes should be statistically equivalent.

3.2 Issues concerning computer-adaptive/assisted language testing (CALT)

The use of computerized adaptive testing has raised important concerns in testing. Brown (1997) pointed out that in designing a computerized adaptive test a large number of examinees are required in order to pilot the test items and create an item pool based on the IRT methods. Methods to score the computerized tests as well as decisions about the cut-off points are not clear. With regard to logistical concerns, computerized test item disclosure can have some effects on the process of item banking and test security issues.

4. Computerized test delivery systems

From the previous section, it seems that the advantages of using computers in testing outnumber the drawbacks. Computer technology has been substantially improved to suit the testing situation. The computer will be propitious and widely applied for assessment in the future. There are many aspects that make computerized tests differ from conventional ones. Computerized tests are presented on a computer screen in which the examinees have to correspond with the test by clicking, typing, or touching the screen directly. Besides the distinctive test presentation, computerized tests also differ from one another in terms of test delivery processes. There are several test delivery methods that are in practice. These methods are distinct and can be used to suit the requirements of different test administrations. Computerized test delivery systems are composed of “hardware, software, testware, and human expertise necessary to deliver the intended instruction or measurement” (Bunderson et al, 1989: 370). Four major test delivery methods are reviewed below.

4.1 Computerized fixed tests

This method can be viewed as a computerized version of a paper-and-pencil test. The paper-and-pencil test is converted to “a computer-administered format” (Bunderson et al, 1989). It is the least sophisticated test delivery method. It delivers the test in a linear fashion (Parshall et al, 2002). This method does not require sophisticated algorithms and psychometric models. It also does not require very complicated computer components to deliver the test (Bunderson et al, 1989). A test delivered through this method has a fixed length and a fixed test form. In the fixed test form, items can be administered in a fixed order or randomly. Test forms can be constructed based on classical item statistics and equated by conventional procedures.

Scores are given for items answered correctly and they can be reported immediately after the test completion or later (Parshall et al, 2002). This test delivering method also enables standardization of the test since it has a preset format for test presentation. As a result, it can increase test security (Bunderson et al, 1989).

4.2 Automated test assembly for computer delivery

Parshall et al (2002) described this delivery method as a hybrid between the first simple delivery process and an adaptive test system. In this delivery method, test items are collected and saved in an item pool. The items will be selected and they will form different test forms. These items are constructed and analyzed for equivalence under certain criteria as specified in the test development procedures. The items are constructed and developed using classical item statistics. Hence, they are not adaptive. Once many forms of the test items are arranged, each form will have a fixed length. The items in the item pool can be restored or revised after being overexposed for some time. The tests can be “scored by classical test theory methods or estimated for each examinee’s ability as measured by the test” (p. 11). That is traditional test scoring methods can be applied. The scores can be reported after the test completion or later.

4.3 Computerized adaptive testing

The adaptive test system can be explained as follows: “the presentation of the next task (test items) depends on calculations based on the test taker’s performance on previous tasks or items” (Bunderson et al, 1989: 381). Parshall et al (2002: 11) stated that “the goal of this type of test is to obtain a precise and accurate estimate of each examinee’s proficiency on some underlying scale.” The test takers will not experience the same items and will have received an unequal number of test items

when completed. The test items are analyzed using the item response theory (IRT) parameters. IRT presupposes that test takers “differ in their ability on a unidimensional continuum ranging from low to high ability. For each examinee, the probability of answering each item correctly depends on the current ability estimate of the examinee and the properties of the item response curve for that item. Item response curves are usually specified up to three parameters: the location of their most effective point (the difficulty), their slope at that point (the discrimination), and their intercept (guessing parameter)” (Bunderson et al, 1989: 382). As a result of the IRT parameters there are fewer test items to administer and the scores gained are more reliable and precise. Since the examinees correspond to the unique test item set estimated for their ability in real time, the scores are to be equated “through reliance on IRT ability estimates” (Parshall et al, 2002: 126).

4.4 Computerized classification tests

Computerized classification tests (CCT) are similar to the computerized adaptive tests in that they are also adaptive. Computerized classification tests are useful for continuous assessment and are normally used to classify the examinees into groups such as “pass” or “fail”. Scores may not be reported. Even though the selection of test items and the item information can be done based on IRT parameters or other similar statistics approaches, the information regarding the ability estimation of the examinees may not be fully provided (Parshall et al, 2002: 153-168).

For this study, the computer program that was used to deliver the reading comprehension test was developed based on the principles of the Computerized Fixed Test delivery system mentioned above because the study only aimed to compare test performances on the two modes. In other words, the study was carried out to

investigate effects from modes of test presentation on the test performance. Therefore, the Computerized Fixed Test delivery system was the most suitable to be used for this study.

In short, the computer is now a potential device in assessment. At present, there are four delivery systems that are commonly used in assessment. The computer can be exploited by both test takers and testers to measure the test takers' true abilities. However, there are still issues and concerns about the use of computer in assessment. To unravel some issues related to the use of the computer in assessment such as validity of the computerized tests and the equivalence between the computerized and the paper-and-pencil tests, some research studies investigating those issues are reviewed in the following section.

Studies related to comparing conventional and computerized tests

1. Equivalence of the conventional and the computer-based tests

Zandvliet and Farragher (1997: 423-38) designed a computer-testing program using HyperCard software in order to investigate the equivalence of "computer-administered tests (CAT)" and conventional tests and to compare and investigate the attitudes of the students towards the CAT and the conventional test. The subjects in this study were 50 adult basic education students in an introductory computer course. They were randomly divided into two groups. There were three computer-administered tests and three written tests. The students had to complete all the tests. The students were asked to fill in the surveys that were distributed before and after taking the computerized tests. The pre and post questionnaires were similar. Their aims were to find out the students' attitudes about computers, computerized testing,

and conventional testing. The statistics used to analyze the data were mean and t-test. The results of the comparative analysis of tests scores indicated that there were no significant differences between computer-based and written test scores. The responses from the questionnaires were reported in the form of percentages. The prior survey responses revealed that the students had a preference for the computer-based tests over the written ones and that this preference increased after they took the computer-based tests. The researchers also studied the students' test-path and found that the students spent, on average, two minutes longer to complete the computerized tests.

Olsen et al (1986 cited in Bugbee, 1996) conducted a study on the equivalence and equation of three methods of test administration—paper administration, computer administration, and computer-adaptive testing. The tests in this study were achievement tests of mathematics application items taken from the California Assessment Program. The subjects were 350 third graders and 225 sixth graders. The students at each grade level were randomly divided into four groups using a counter-balanced design. Each group received two different versions of the tests. The results revealed that there was no significant difference between paper-administered tests and computer-administered ones.

Choi et al (2003: 295-320) conducted a study on comparability of a paper-based language test and a computer-based language test in the context of the English language as a foreign language. The participants in this study were students from five universities in Korea. They were randomly divided into two groups and they had to take both forms of the tests: computer-based and paper-based tests. The instrument in this study was derived from one test of 200 items—Test of English Proficiency

(TEPS) developed by Seoul National University. The TEPS was designed to measure four areas: listening comprehension skills, grammar, vocabulary, and reading comprehension skills. Half of the test was administered as a computer-based test (i.e. 100 items) and the rest (i.e. 100 items) was a paper-based test. To find the comparability between the two forms of the test, particularly on content and construct validity, the researchers employed content analyses based on corpus linguistic techniques, as well as statistical analyses: correlational analysis, ANOVA, and confirmatory factor analysis. They concluded from their findings that the two subtests of TEPS—computer-based language test and the paper-based language test—were comparable in terms of content and linguistic features. “The overall results of construct-related validation studies indicated comparability of the subjects’ scores across the computer-based language test (CBLT) and the paper-based language test (PBLT). The grammar test showed the strongest comparability and the reading comprehension test the weakest comparability. The pattern of correlations among subtests, disattenuated correlations, and confirmatory factor analyses supported to a certain extent that CBLT and PBLT subtests measured the same constructs (p. 316).”

Sawaki (2001) presented a review of the literature and research studies on computerized testing. Based on research studies, there are many important issues related to the computerized and paper-pencil testing that should be considered. Among those issues is an impact on examinees of the introduction of computerized tests. Sawaki pointed out some results from a study conducted by Taylor et al (1998) on the effects of computer familiarity on examinees’ performance. In the study on the CBT version of the TOEFL, the examinees were provided with computer training and a paper-and-pencil TOEFL in order to adjust their ability levels. Then, the examinees

were assigned to two groups: computer familiar and computer unfamiliar. The researchers found that there was a small statistically significant difference. However, before the adjustment for ability, differences on the performances within the familiarity groups were statistically significant. But after adjustments were made for the level of ability measured by the paper-pencil TOEFL, there was no statistically significant difference between computer familiarity and test performances. “Thus, the researchers (Taylor et al) concluded that computer familiarity does not play a major role in CBT TOEFL performance” (Sawaki, 2001: 42).

2. Studies on mode effects

Clariana and Wallace (2002) conducted a study to find out whether certain factors—namely, content familiarity, computer familiarity, competitiveness, and gender—influence test performances on paper-and-pencil tests and computer-based tests. The participants in the study were two groups of students. One group was specified to take the paper-and-pencil test and the other group the computer-based test. The test on both modes was about the subject they studied. They found that the participants taking the computer-based test outperformed the participants taking the paper-and-pencil test. This difference was a result of content familiarity but not other factors (gender, competitiveness, and computer familiarity). They stated that the computer-based tests helped the high-attaining students more than the paper-and-pencil tests.

Dillon (1992: 1297-1326) reviewed empirical literature about issues of reading from electronic texts and from screens in particular. He stated that reading behavior, especially from screens, was assessed based on the outcome of the reading and reading process. The “outcome measures concentrate on what the reader gets from

the text” and consider variables such as amount of information retrieved, accuracy of recall, time taken to read a text, fatigue, comprehension, and much more. The “process measures” concentrated on “how the reader uses a text including variables such as eye movements, manipulation, and navigation” (p. 1299). Since a large number of studies on those issues in reading from screens were summarized in this article, only the studies related to the present study will be reviewed.

Some studies that were conducted in the 1980s investigated the effects of presentation medium on comprehension from the perspective of reading outcome. These studies required the subjects to read from screens and respond to comprehension questions after the reading tasks. Two studies by Muter et al (1982) and Kak (1981) cited in Dillon (1992: 1303) reported similar results. It was found that there were no effects from the presentation medium on comprehension. Belmore (1985 cited in Dillon, 1992: 1303) investigated reading comprehension and speed spent in reading from both printed materials and from screens. Belmore found there were some difficulties in comprehension and speed in reading from screens. Further investigation revealed that decrement of the performance on the tasks presented on the screen was due to “the lack of familiarity with computers and reading from screens” (p. 1303). Another study was conducted by Muter and Maurutto (1991 cited in Dillon, 1992: 1303-1304). They compared reading comprehension from print and screens and found that there were no significant differences between reading comprehension performances from the two media. Regarding the subjects’ preference to reading from screens, Muter et al (1982) mentioned that the subjects tended to prefer reading from a book to a screen. This was because the subjects thought that the books allowed them to turn back pages and re-read but the computer did not allow

them to do so. Muter and Maurutto (1991) investigated further into the subjects' preferences regarding reading from screens and from paper. They found that their subjects showed a preference for reading from screens. Thus, Dillon concluded that the alteration of preference from reading from paper to screens was due to the improvement of screen technology.

3. Students' attitudes towards the computer-based materials

Coniam (1999) conducted a study to investigate subjects' reactions to computer-based tests composed of different format types: multiple-choice, proofing-cloze, and text dictation. The subjects were Hong Kong trainee teachers whom the researcher claimed were experienced computer users. The instruments employed in the study were computer-based tests and a post-test questionnaire. The results showed that the subjects had ambivalent reactions to the computer-based tests. Responses to the questionnaire questions about the subjects' preference, readiness, and enjoyment taking the tests were not as positive as expected. The researcher suggested that the subjects should understand what to do in a computer-based test and should be able to complete it the same way they do with the conventional test. If this does not happen, the reliability of the test may be affected.

The results of this study suggest that the subjects doing computer-based tests should be familiar with the formats of the tests. In other words, the formats of the computer-based tests should be similar to those of the conventional tests in order to prevent the effects of unfamiliarity of the test formats from interfering with the study.

Ketkham (2004: 67-94) studied students' attitudes toward "computer usage and the content of the computer-based reading program (CBR)" and investigated whether their attitudes affected their reading performances. The subjects of the study

were 230 graduate students who were asked to do reading activities on the computer. They were also asked to complete a self-report questionnaire at the end of the course. The results revealed that the students showed positive attitudes toward the computerized reading activities. Moreover, the attitudes towards the use of the computer contributed to the reading performances while the attitudes towards the content of the program had no influence on the performances. The researcher stated that the attitudes towards the use of the computer could be a predictor for the reading performances.

In short, the results from the reviewed studies investigating the equivalence of the computer-administered and paper-administered tests revealed similar results. The findings from those studies indicated that tests presented on paper and on computer screens were equivalent and comparable. It seems that the two modes of testing do not have an impact on test performance. In terms of the subjects' attitudes or preferences towards the two modes of test presentation, the subjects in those studies revealed positive attitudes or preference towards the computerized tests. However, in some studies, the subjects stated that they preferred taking the tests presented on paper and reading from books because of their familiarity with the conventional way of reading and testing. The effects of familiarity with computers can be controlled and minimized in the developmental stage of computerized tests as suggested in the studies reviewed by Dillon (1992). However, there are still a number of considerations regarding the properties of the computer that can affect test taking processes and performances.

Summary

Reading models in general, reading comprehension, and models of reading in the second or foreign language are reviewed in this chapter because they are related to reading processes. The implications and uses of the computer in language assessment are then described. It is apparent that the computer will be widely used in the field of assessment in both high-stakes and low-stakes situations. Research studies on the use of computers in assessment are also presented. The results of those studies are similar in that computerized tests can be equivalent to the conventional ones. Moreover, the test takers showed positive attitudes towards the computerized tests. This implies that the computer can be a test-taker friendly test format. The research designs in some studies that are reviewed in this chapter are applied in the present study and presented in the subsequent chapters.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER III

METHODOLOGY AND PROCEDURES

This chapter explains the research methodology, population and sample of the study. The development of the Reading Comprehension Test of English (RCTE) and a questionnaire, and the data collection procedures are also demonstrated in detail. Included in the chapter are results of the instrument validation.

Research design

This study is a quasi-experimental research project. The treatments are two different modes of test administration: (1) a reading comprehension test of English (RCTE) presented in the paper-and-pencil mode and (2) a reading comprehension test of English (RCTE) delivered through the computer. The self-control subjects of the study were given the two modes of test administration.

Population and sample of the study

1. The population

The population in this study were first year students from the Faculty of Humanities enrolled in the academic year of 2004 at Chiang Mai University (CMU). Four hundred and seventy of them were studying Foundation English (FE I, II, and III or English 103, 104, and 203, respectively). They were aged between 18 and 19 years. Most of them had graduated from secondary schools in the northern region of Thailand, though some were from other regions in the country. During their secondary school years, the students who were enrolled in the Faculty of Humanities

at CMU concentrated on foreign languages and social science studies. Based on their scores from the English language test on the Entrance Exam (EE) or the Northern Quota Exam, they were assigned to one of the FE courses. The students with an English score from the EE ranging from 0 to 70 were studying FE I. The students in the FE I were divided into two subgroups: FE I Group I and Group II. FE I Group I included students who obtained English test scores from the EE of 0-40. FE I Group II consisted of the students whose EE scores ranged from 41 to 70. Those with EE scores ranging from 71 to 80 registered for FE II and those who gained scores of 81 or higher were in FE III. Of all students from the Faculty of Humanities, 45 of them were in FE I group I and 335 students were in FE I group II. Sixty four students were studying FE II and only 26 students were taking FE III. Even though they appeared to be relatively heterogeneous in terms of the English courses they had taken and their educational backgrounds, they were currently studying in the same faculty and taking the same humanistic courses.

For this study, English language proficiency levels based on the EE score were rearranged. Those proficiency levels are low, intermediate, and high and the English test scores of each level were 0-50, 51-80, and 80 or higher, respectively. The students were then reclassified into those three levels of English proficiency.

2. Sample

The sample in this study was an intact group. They were purposefully selected from certain sections of the FE I, II, and III and 66 students participated. They were drawn from all English proficiency levels based on their EE scores. Among the 66 students, 11 were male students and 55 were female students. Of 66 students, 23 (34.85%) students were from the low level, 35 (53.03%) students were from the

intermediate level, and 8 (12.12%) students were from the high proficiency level. The proportion of the students in each proficiency level and their majors are illustrated in Tables 2 and 3 below.

Table 2. Number of the Students Participating in the Study

EE scores	0-50	51-80	81 and higher	Total
Number of students	23 (34.85%)	35 (53.03%)	8 (12.12%)	66

Most were majoring in French, Mass Communications, and Home and Community. The rest were from English, Philosophy, Thai, Japanese, and Chinese majors as shown in Table 3.

Table 3. Majors of the Students Participating in the Study

Majors	Number of the students	Percentage
French	22	33.33
Mass Communications	15	22.72
Home and Community	12	18.18
English	8	12.12
Thai	6	9.09
Philosophy	1	1.512
Japanese	1	1.512
Chinese	1	1.512
Total	66	100

In addition, another 35 students from the intermediate level (i.e. their EE scores on the English test section ranging from 50 to 70) participated in the pilot study to validate the instruments. They were randomized using the Stratified Random Sampling technique. Because this study was not affiliated with any particular English

courses offered at CMU, participating in this study was not mandatory for the students. As a result, the random plan in the pilot study was not successful. Because of this problem with the random technique, the sampling plan was changed to the intact group option.

Instruments of the study

The instruments used in this study were the Reading Comprehension Test of English (RCTE), an attitude questionnaire on modes of testing, and a computer program to deliver the computer-based RCTE. Their development processes are explained in details below.

1. Development of the Reading Comprehension Test of English (RCTE)

The test development framework by Bachman and Palmer (1996) was applied in the test designing stage for this study.

1.1 Defining the test constructs

From the models and issues of SL/FL reading reviewed in the previous chapter, it is apparent that schemata and background knowledge play a significant role in the reading processes. They can affect test performances and distort the true ability that is being measured (Bachman and Palmer, 1996). In a test development process, test developers should be aware of schemata effects. Bachman and Palmer (1996) propose three approaches to prevent and circumvent constraints that can be caused by background knowledge in the test designing stage. One option is that the background knowledge is excluded solely from the constructs. In other words, only the language ability that we want to measure is defined. Another way is to include background

knowledge and language ability in the constructs. The last possibility is to define background knowledge as a separate construct from the language ability.

For this particular study, the background knowledge or schemata were excluded from the constructs. Only the reading comprehension skills in the SL/FL were measured. In addition to preventing the effects of background knowledge, factors related to the students and their English courses taken at CMU were also considered and minimized in the test development process. Since the English courses offered to all first year students in the academic year of 2004 are absolutely new, passages and vocabulary used in the tests had to be neutral and accessible to all of the participants.

A Foundation English II course syllabus offered to first year students at Chulalongkorn University Language Institute (CULI) was selected as a guideline for the test construction. A set of vocabulary and two themes for the reading passages were chosen: technology and alternative therapy.

1.2 Design statement of the RCTE

1.2.1 Purposes of the test

The RCTE was developed to measure the general English reading comprehension skills of first year students in the tertiary level and to find out whether two different modes of test presentation (paper-and-pencil and the computer) affect test performances.

1.2.2 Description of the target language use and task types

The target language use in this study refers to the ability to apply reading skills when approaching reading materials from real-life magazines, newspapers and

textbooks which contain articles or passages written in English and in different genres such as narrative, descriptive, explorative, and argumentative.

The task type of this test was in the multiple-choice format (MC). Although the MC format has been increasingly criticized, it is still being employed in many test constructions. Alderson et al (1995: 45) pointed out that the MC format is often used in testing reading ability since it allows test writers to “control the thought processes of readers.” Popham (1978: 55) also stated that the MC format is “flexible” and “can be applied to measure cognitive and affective outcomes.” It is also objectively scored which can create reliability. However, there are some weaknesses in the MC format that were also considered during the test writing stage. To make this RCTE a reliable one, the empirical data underwent item analysis. The results were used to revise the test items. The RCTE is composed of two tasks: vocabulary in context and reading comprehension.

Task One: Vocabulary in context. This task is comprised of a stem, which is a short piece of text with a blank for a missing word. The missing word is to be a noun, a verb, an adjective, or an adverb. Four different words are provided in the four-option MC format. One of those is the correct and the most appropriate word to complete the statement. The others are distractors.

Task Two: Reading Comprehension. This task is composed of a reading passage of approximately 450-500 words, which is 3-5 paragraphs. There are test items following the passage. A test item is comprised of a stem which can be a direct question or an incomplete statement. Four answers are provided in the MC format. One of them is the correct answer and the rest are distractors.

1.2.3 Characteristics of test takers

The test takers were 18 and 19 years of age. They were admitted to study in the Faculty of Humanities at Chiang Mai University (CMU). They may have different levels of English proficiency because of their exposure to the English language while they were in secondary school. They were expected to be familiar with the four-option multiple-choice testing format since they had taken either the Entrance Exam or a quota exam which were in such a test format. With their educational backgrounds, many of them studied in foreign language stream, social science, and humanistic programs during high school. At the time of the implementation of these tests, the test takers were majoring in one of the humanities fields, such as English, French, German, Japanese, Thai, and history. The tests used in this study should not create anxiety among the students/test takers since they were specially designed without reference to any specific English language courses offered at CMU. The scores gained from the RCTE were not used in any grading decisions. In other words, the test of this study could be viewed simply as a low-stakes test.

1.2.4 Operational definition of the test construct

Reading comprehension in the present study means “the ability to understand information” in a given reading passage and “interpret it appropriately” (Grabe and Stoller, 2002: 17). The tests were constructed based on a syllabus of FE II offered at CULI and a theory of reading in the second/foreign language. Background knowledge or schemata were not assessed. The tests focus mainly on drawing out reading abilities and the ability to understand text organization or knowledge of organization (Bachman and Palmer, 1996) and knowledge of the English language. To be able to achieve comprehension, the test takers or the participants of this study

should be able to exhibit the following reading skills and knowledge (i.e. knowledge of English language).

Reading comprehension skills

- Scanning
- Skimming
- Identifying pronoun references
- Guessing word meaning from context
- Identifying the topic and main idea
- Recognizing supporting details
- Summarizing information
- Making inferences
- Drawing conclusions

Knowledge of the English language and organization

Knowledge of the English language is the knowledge of organization that, in this situation, emphasizes only knowledge of grammar (comprising vocabulary and syntax), textual knowledge (including cohesion) and pragmatic knowledge.

1.2.5 Plan for evaluating the qualities of usefulness

The qualities of usefulness of the test are composed of reliability, construct validity, authenticity, interactiveness, impact, and practicality. The plan to achieve these qualities will be outlined below.

Reliability

This test was a low-stakes test due to the fact that the scores that the participants gained did not affect their performances in their current English courses. However, the K-R20 reliability level was set at a high level of .80. The reliability was obtained through the trying-out, pilot study, and the main study.

Construct validity

An important issue in designing and using any language tests is that the test results should enable the test developers to make inferences about the performance of the test takers beyond the test situations. The tests must also elicit the test takers' true language abilities that the tests are aimed to measure (Bachman, 1990). To make the RCTE valid and its results purposeful, it was constructed based on theories of SL/FL reading comprehension, particularly the interactive models of reading. Three specialists were asked to validate the tests. The tests were then tried out in order to do the item analysis.

Authenticity

The selected reading passages were assembled from those in real life and they were not simplified. Simplification of a text can be done in several ways such as simplifying the syntax features or changing lexical items. Though simplification may make the text less difficult, it may distort the original message or may make the text more difficult in some other aspects of text features (Alderson, 2000a: 72-74). However, some parts in certain passages were discarded to make it suitable for the test situation. In real life it is quite unusual to add questions or quizzes about every passage or articles for the readers to answer like in a test situation. However, the readers themselves may form questions about a passage they read.

Interactiveness

The interactiveness of this RCTE refers to the interactive reading process in which the test takers apply both low-level knowledge and high-level knowledge while reading the text. A way to observe this interactiveness is to include test items which

measure both low-level processing components such as word recognition and high-level processing components such as syntax and semantic knowledge.

Impact

The implementation of the RCTE may not have serious impact on the test takers or the participants since it is not related to their studies and grading. However, the participants may feel more conscious about their reading abilities. The results from the RCTE delivered through the computer would somehow initiate considerations among practitioners in educational organizations about implementing the computer in instruction. The impact of the implementation of a computerized test can be obtained by interviewing the students or test takers and individuals involved in the test development and administration.

Practicality

The issues related to the use of the RCTE especially through the computer mode are very important and need thorough consideration. Even though a computerized test can assist the test developers and the test takers in many ways, the management of computerized tests seems to be quite complex and requires quite a number of staff, especially technicians and computer programmers. Besides, the staff who administer computerized tests are required to have knowledge and skills of the systems.

1.3 The test blueprint

1.3.1 Test structure

Number of parts/tasks: The RCTE is composed of two main parts: (1) vocabulary and (2) reading comprehension.

Sequence of parts: The RCTE begins with the vocabulary part and it is followed by the reading comprehension part, which is divided into two subsections: reading comprehension I and II.

Number of tasks per part: The vocabulary part consists of 40 test items and the reading comprehension section consists of four reading passages. Each reading passage is about 450-500 words long and is followed by 15 test items. Hence, the reading comprehension section is composed of 60 test items. The total number of test items is 100. These 100 items are divided equally into two halves which are presented in the two modes of testing. This means 20 vocabulary items and two reading passages with 30 test items are presented on paper. The other set is presented on the computer, that is, the other 50 test items are computerized.

1.3.2 Test task specifications

Purpose: The RCTE was developed to measure reading comprehension of English from two modes of test presentation: conventional (paper-and-pencil) and the computer.

Definition of the construct: The construct was the same as in the design statement.

Characteristics of the setting: The setting of the RCTE administration consisted of two locations. The paper-and-pencil test was administered in a classroom in which seats were arranged in a testing situation and the computer-based RCTE was held in a computer lab, which consisted of 42 desktop sets. Each computer was set apart from the others without cubicles. Obviously, the computer lab was set up this way to facilitate computer instructions and not for testing purposes.

Both locations were in the same building but on separate floors at the Faculty of Humanities at CMU. The rooms were well-lit and air-conditioned.

Participants: The participants were first year students from the Faculty of Humanities who were, at that time, studying Foundation English I, II, and III (FE I, II, and III) at CMU. Besides the test takers, the researcher attended and monitored the computer-based test while two research assistants monitored the paper-and-pencil test.

Time of task: Since this study was specially designed and was not affiliated with any English language courses offered at CMU, and in order not to disturb the participants' regular class periods, the RCTE had to be administered out of the regular class times. Fortunately, under a new regulation at CMU, there are no classes on Fridays. Therefore, the study was held on Fridays during the first semester of the academic year of 2004 at CMU. The test administrations were arranged to meet the participants' convenience.

Time allotment: In the pilot study, each mode of testing took about 2 hours. For the main study the time allotment for each mode was reduced to 1.30 hours due to the reduction of the test items.

Instructions: The instructions for each mode were in Thai.

Characteristics of input:

Vocabulary

General Description: After reading a short text with a blank, the test takers had to select an appropriate word from the given four multiple-choice options.

Prompt attributes:

1. Each short text was written up or adapted from articles or short passages related to two topic areas: technology and alternative therapies. They were taken from newspapers, textbooks and magazines.
2. Knowledge to be tested was knowledge of the English language:
 - 2.1 Grammatical knowledge: parts of speech, collocation, appropriate meaning
 - 2.2 Abilities to use contextual clues
3. Each text was consisted of about 3-5 sentences long with a blank of a missing word. The word was one of these: a noun, a verb, an adjective or an adverb.
4. Each text was made up of facts, descriptions, arguments, or a combination of these.

Response attributes:

1. On the paper-pencil test, the test takers selected and circled an appropriate word from the four multiple-choice options. For the computer-based test, the students used a mouse to click the option of their selected answer.
2. The four multiple-choice options comprised:
 - a. One correct answer that was grammatically correct and had a meaning that appropriately completed the sentence or the text.
 - b. Three incorrect choices that were grammatically correct but had an inappropriate meaning or were both grammatically incorrect and had an inappropriate meaning.

Reading Comprehension

General description: After reading a passage of 450-500 words, the test takers were to answer comprehension questions based on the passage.

Prompt attributes:

There were four reading passages.

1. Each was related to one of two topic areas, technology and alternative therapies, and was taken from a textbook, a newspaper, or a magazine.

2. Each contained 450-500 words.

3. Each passage included facts, description, argumentation, report, or a combination of these.

4. Each passage was checked for readability level. The readability level expected for each passage was at the 10-12 Grade level using Flesch-Kincaid Grade Level readability formula available in the Microsoft Word Processor program.

5. There were 15 test items asking about each passage in the pilot study and the total number of items for all passages in the main study was reduced to 24. The test items were:

5.1 Questions asking the test takers to do one of the following:

- a. find the main idea of a paragraph, understand an implied main idea, choose the best title
- b. find specific information (scanning)
- c. identify pronoun references
- d. guess the word meaning from the context

- e. recognize supporting detail(s)
- f. make inference(s) or draw conclusions

Response attributes:

1. On the paper-pencil test, the test takers had to select and circle the best answer from the four multiple-choice options. On the computer-based test, the students used a mouse to click the correct answer.

2. The four multiple-choice options included

2.1 The correct answer which was one of the following:

- a. general information/main idea of the paragraph or of the passage
- b. specific information
- c. a noun or phrase that is referred to
- d. a synonym or word/expression of similar meaning
- e. a supporting detail
- f. an appropriate inference or interpretation

2.2 Three distractors.

Scoring method: Each correct answer was worth one point. The total score for the pilot study was 100 (i.e. 50 points for the paper-and-pencil test and 50 points for the computer-based test). For the main study, the scores were reduced to 44 points (i.e. 22 points for the paper-and-pencil test and 22 points for the computer-based test).

2. The development of the questionnaire

2.1 Purpose

The questionnaire was developed to find out the participants' attitudes towards the two modes of testing. Specifically, the questionnaire aimed to compare the participants' attitudes/opinions about the two modes of test presentation before and after they took the tests in the two modes.

2.2 Constructs and organization

The questionnaire was in the form of a five-point Likert's scale. The scale of one represents disagreement or the least preference and the scale of five represents strong agreement or the most preference. The questionnaire was composed of four main parts.

Part I: English language test anxiety. This part examined the participants' anxiety when taking the English language test in general. This part was composed of nine items.

Part II: Familiarity with the computer. This part asked about the participants' familiarity with computer use and computerized tests. There were nine items.

Parts I and II were adapted from a study of Stricker and Wilder (2001) on *Examinees' attitudes about the TOEFL-CBT, possible determinants, and relationships with test performance* (TOEFL Research Report 01-01).

Part III: Attitudes towards conventional and computer-based tests. This part was developed by the researcher in order to find out the participants' attitudes/preferences about tests presented in two different modes of testing (conventional and computer) before and after they took the tests. This section of 17

items was tried out in the pilot study and revised. After the revision, 26 items were tried out again before being implemented in the main study.

The constructs of Part III are outlined below:

Items No. 1-2: Students' familiarity with and anxiety about the tests delivered by the two modes.

Items No. 3- 8: The English language test characteristics: test format, length of the reading passages, difficulty of the test, the language of the instructions, and the font size.

Item No. 9: Students' preference for test presentation modes.

Item No. 10: Time allowance for the two modes of test presentation.

Part IV: Open-ended section. This last part asked the participants to express their preferences and concerns about the two test presentation modes.

3. The development of the computer program to deliver the RCTE

The computer program used to deliver the RCTE was specially designed for this study. The program was written using Visual Basic Version 6.0 and the Database software. The delivery system was in a linear fashion. It followed the Computerized Fixed Test Delivery Method (Parshall et al, 2002) and was composed of two main programs: the test presentation program and the record program (the database). The test presentation program was divided into four main sections: introduction, tutorials

(compulsory), the tests, and the report. The delivery process is illustrated in Figure 1 below. The database (as shown in Figure 2 on page 72) was also divided into two main sections: student's information record and the database of the test which allows the test developer to add an unlimited number of test items and revise the existing items.

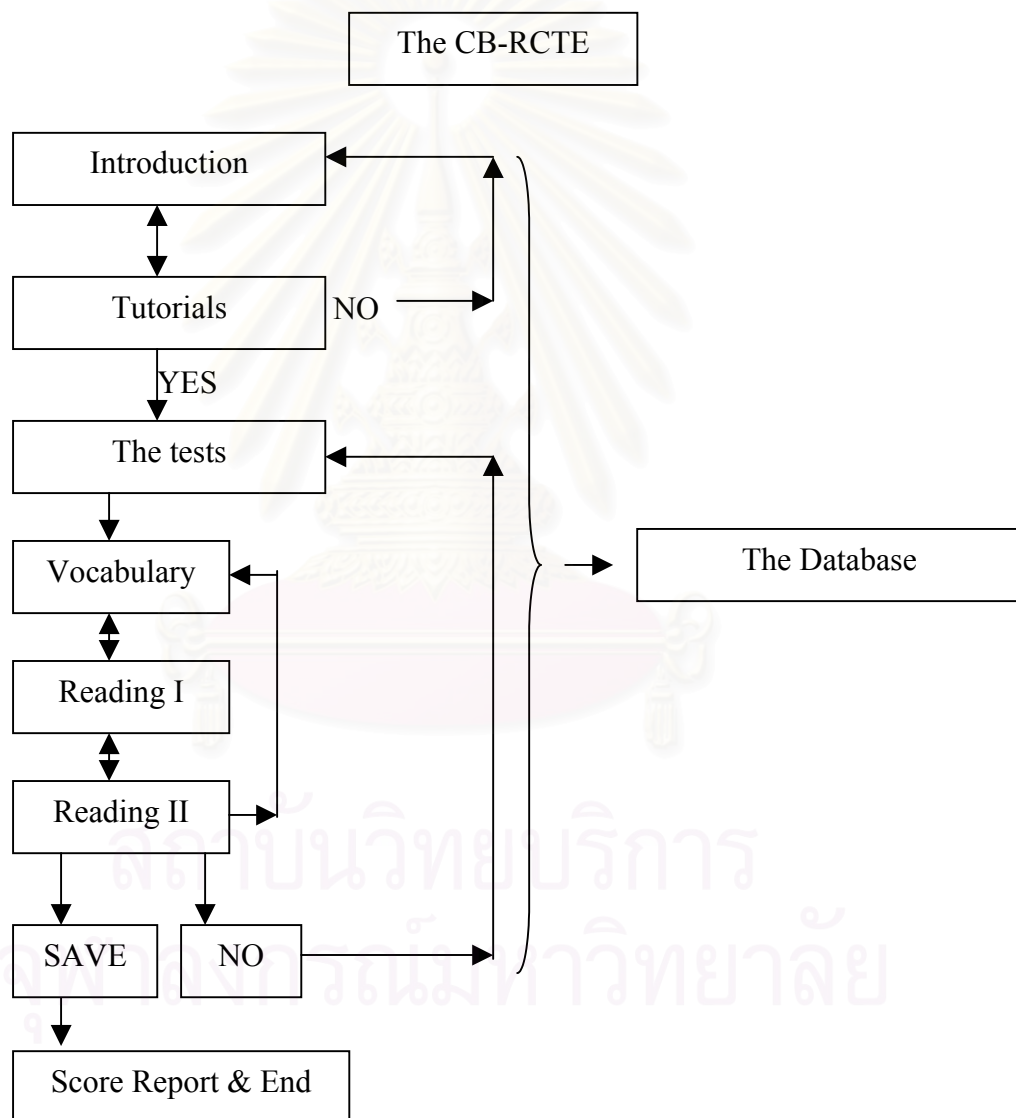


Figure 1. The Program Outline of the Computer-Based RCTE

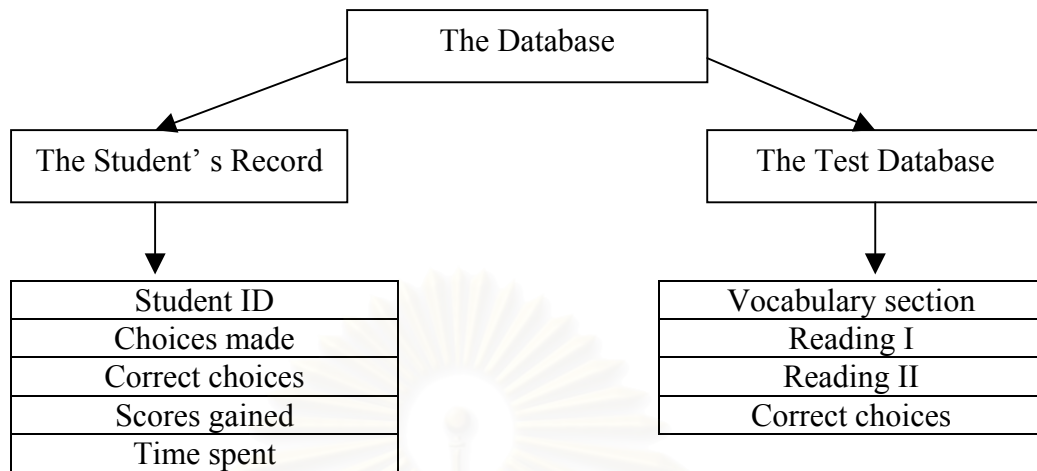


Figure 2. The Database of the Computer-Based RCTE

The Tutorials

The tutorials were compulsory in order to familiarize the test takers with the CB-RCTE system. The test takers had to complete the tutorials before navigating to the test sections. The tutorials were divided into three sections: (1) how to use a “mouse”, (2) how to scroll the text, and (3) how to choose the answer. These activities were arranged in a fixed sequence and the students could navigate to subsequent sections when they completed the preceding one. The time allowed for the tutorials was 20 minutes. However, the students could spend less than the time allowed in completing the tasks. This meant they could begin the test immediately when they finished the tutorial section.

Procedures

1. Development of the instruments

The instruments, namely the RCTE, the questionnaire, and the computer program to deliver the RCTE, were developed and revised during the second semester of the academic year of 2003 (during October – February, 2003). The instruments were developed according to the specifications described in the previous section. The RCTE was initially composed of 40 vocabulary items to measure the ability to use words in contexts and four reading passages with 60 test items. The questionnaire was composed of four parts and the total number of items was 35.

2. The instrument validation

The RCTE and the questionnaire were examined for content validity, construct validity, and criterion-related validity. The content of the RCTE was based on the objectives for reading ability of the FE II offered at Chulalongkorn University Language Institute (CULI). The reading construct of the RCTE was based on the interactive model of reading in SL/FL. Apart from the content and construct of reading, a simplified version of Bloom's Taxonomy of Educational Objectives that is used in test designing was applied. The criterion-related validity was obtained by using the Pearson Product Moment Correlation Coefficient to find the correlation coefficient between the RCTE and the English language test scores from the Entrance Exam. The reliability level of the two instruments was set at .70 (K-R20). The validation process and results are described below.

2.1 A priori validation

2.1.1 Qualitative analysis of the RCTE

The RCTE was originally composed of 40 vocabulary test items and four reading passages with 60 test items (i.e. 15 items corresponding to each passage). Thus, the total number of the test items was 100. The RCTE was developed according to the specifications mentioned in Chapter III. The vocabulary items were created to measure the ability to use words in context and knowledge of the English use and grammar. All four reading passages were examined for the readability levels using the Flesch Readability formula available on the Microsoft Word Processor. Two values were used in consideration: grade level and reading ease level. According to the Flesch formula, the higher the grade level the more difficult the text is. The lower the reading ease level the more difficult the text is (Flesch, 1951).

The reading passages selected were in the areas of technology and Internet and alternative therapies. There were two reading passages under the theme of technology, namely “Online Epidemic” and “No Calls in Class, Kids”. The “Online Epidemic” passage had a level of reading ease at 38.5 and its readability level was at Grade 12. The “No Calls in Class, Kids” passage had a level of reading ease of 53.2 and its readability level was at Grade 10.7. The other two reading passages under the alternative therapies theme were “Meditation” and “Naturopathy”. The passage “Meditation” showed a level of reading ease at 47.5 and its readability level was Grade 11.1 while the “Naturopathy” passage revealed a reading ease level of 39.1 and its readability level was at Grade 12. The readability levels of the passages are shown in Appendix C. The four reading passages could be considered relatively difficult and

should be appropriate to the level of English proficiency of the participants in the study who were studying in the first year at the tertiary level.

Three lecturers at the Chulalongkorn University Language Institute (CULI) examined the RCTE's construct and content appropriateness. The contents and reading skills to be measured were compared against the reading skills in second/foreign language theory and in Bloom's taxonomy of the cognitive domain. Based on Bloom's taxonomy of the cognitive domain (Bloom, 1974), there were sixteen test items that aimed to measure the knowledge level. Nine items measured skills at the comprehension level while eight items aimed to measure skills at the analysis level and 27 items for the synthesis level. Looking at specific reading skills, there were six items testing the ability to find pronoun references, ten items on finding specific information, and nine items measuring understanding of both explicit and implied ideas. The abilities to make inferences and to find main ideas were measured in eighteen items. Seventeen items aimed to measure the ability to guess meaning from the context while only one item measured the ability to find a topic or a title. Therefore, the total number of test items in the reading comprehension section was 60. The proportion of the test items and reading skills measured is illustrated in Appendix C. For the vocabulary part, all vocabulary items were also examined for content appropriateness.

2.1.2 Qualitative analysis of the questionnaire

The questionnaire was developed according to the specifications mentioned in the previous section of this chapter. There was only one set of the questionnaire and it was composed of four sections. Part 1 consisted of nine items surveying the participants' anxiety towards the English language test in general. Part 2, consisting

of nine items, was to find out the participants' familiarity with the computer. Part 3, comprising seventeen items, aimed to find out the participants' attitudes towards the test delivered through the two modes of testing. The final part (Part 4) was an open-ended section that allowed the participants to add or express their opinions about the tests from the two modes. The last two parts (Parts 3 and 4) were to compare their attitudes before and after they completed the tests. Therefore, Parts 3 and 4 were distributed twice: before and after the administration of the test. To validate the questionnaire qualitatively the same three lecturers from CULI also reviewed the questionnaire, which is in the Thai language. They agreed on the appropriateness of the content and language. The pilot version of the questionnaire is shown in Appendix E.

2.2 A posteriori validation

2.2.1 The RCTE item analysis

The paper-based (P&P) and computer-based RCTE were administered to 35 students at CMU and the data was analyzed using the Item Analysis and Grading Program (CTIA/Grading version 6.30) developed by Associate Professor Suphat Sukamolson, Ph.D. (2000). The program is based on the classical test theory and it is appropriate to perform statistical analysis of the RCTE. The statistic values of all 100 test items derived from the item analysis procedures are shown in Appendix D. The descriptive statistics from the analysis are demonstrated in Table 4 below.

Table 4. Descriptive Statistics of the Paper-Based RCTE and the Computer-Based RCTE in the Pilot Study

Mode of presentation	No. of items	Total score	\bar{X}	SD	K-R 20
P&P	50	50	21.11	7.42	.83
Computer	50	50	24.09	8.61	.87

N = 35

The total number of the test items administered conventionally in the pilot study was 50. Of 50 items, 20 items were in the vocabulary section and 30 items were in the reading comprehension section corresponding to two reading passages (“Online Epidemic” and “Meditation”). The results of the analysis show that the mean score is 21.11 with the standard deviation value of 7.42. The set also yields the internal consistency estimate or the reliability coefficient (K-R20) of .83.

Similar to the paper-based RCTE, the computer-based version underwent the item analysis procedures using the same program. The computer-based RCTE was composed of 50 test items. There were 20 items in the vocabulary section and 30 items in the reading comprehension section. The two reading passages were “No Calls in Class, Kids” and “Naturopathy”. The results as shown in the table above reveal that the mean score is 24.09 with the standard deviation value of 8.61. The set also has an internal consistency estimate (K-R 20) of .87. Although the mean scores are relatively low, the two sets appear to be consistent.

2.2.2 Correlation between the pilot paper-based RCTE, the computer-based RCTE, and the English test scores from the Entrance Examination (EE)

Pearson Product Moment Correlation Coefficient was used in finding the correlation between the two sets of the RCTE from the two modes and between the RCTE sets and the EE scores. The results show that the four subsections of the RCTE (the paper-based vocabulary and reading comprehension sections and the computer-based vocabulary and reading comprehension section) are significantly correlated. The correlation coefficients range from .633 to .769 at the .01 level (two-tailed). In other words, the four subsections yield from a moderate to rather high correlation. (See Table 5 on page 79.)

When examining the internal correlations of the two subsections (vocabulary and reading comprehension) of each mode, moderate correlations are achieved. Of the paper-and-pencil mode, the two subsections are significantly correlated at .711. When examining the computer mode, the vocabulary and the reading comprehension sections are also significantly correlated at the moderately high level. The correlation coefficient is .769.

The correlations of the four subsections across modes were also examined. The paper-based vocabulary is significantly correlated with the computer-based vocabulary at .695 and it is significantly correlated with the computer-based reading comprehension at .739. Another significant relationship was found between the paper-based reading comprehension and the computer-based vocabulary at .633. The paper-based reading comprehension and the computer-based reading comprehension are significantly correlated at .672. Apart from the correlation of the individual parts,

the total scores from the two RCTE sets (i.e. the conventional and the computer modes) yield statistically significant correlation of .785 at the level of .01 (two-tailed).

Table 5. Correlation Matrix of the Paper-Based RCTE, the Computer-Based RCTE, and the Entrance Examination (EE) scores in the Pilot Study

	P&P VC	P&P RC	CB-VC	CB-RC	P&P Total	CBT Total	EE scores
P&P VC	--	.711**	.695**	.739**	.906**	.765**	.719**
P&P RC		--	.633**	.672**	.941**	.696**	.802**
CB-VC			--	.769**	.713**	.908**	.695**
CB-RC				--	.758**	.966**	.742**
P&P RCTE					--	.785**	.826**
CB-RCTE						--	.767**
EE scores							--

N = 35, N of items = 100, ** $p \leq .01$ (2-tailed).

Note: P&P stands for the paper-based test and CB stands for the computer-based test. VC stands for vocabulary and RC stands for reading comprehension.

The subsections of each mode also yield a moderate to high correlation with their respective total scores. The correlation coefficient among them ranges from .696 to .966. It is obvious that the total score of each mode has a high relationship with the scores from its two subparts. However, the correlation of the subsections with the total score from the opposite mode only reveals a moderate correlation. The

relationship between each subpart of each mode also produces a moderate and significant correlation with the English test scores from the Entrance Examination. The correlation coefficients range from .695 to .802.

In addition, the correlations of the total scores of the RCTE from both modes and the English test scores from the Entrance Exam were also calculated. The results show that the paper-based RCTE yields a high and significant correlation with the English scores from the EE at .826 while the computer-based RCTE is significantly correlated with the English scores from the EE at the moderately high level. The correlation coefficient is .767 (.01 level, two-tailed).

2.2.3 The RCTE items selection

Since the RCTE was administered in the two modes and is composed of two separate sets, parallelism between the two sets is essential and should be achieved. In order to claim that the scores of the two sets are equivalent or can be used interchangeably, “theoretical assumptions underlying procedures for establishing score comparability” (American Educational Research Association, 1999: 57) should be provided. For this study, statistic values obtained from the item analysis and the measured reading skills are applied in the selection. Table 6 shows the statistic values that were used as the criteria in selecting appropriate test items for the main study.

Table 6. Statistics Used in Selecting Items

Difficulty Indices:	$0.20 \leq p \leq 0.80$
Delta:	$9.5 \leq \Delta \leq 16.50$
Discrimination Index:	$d \geq 0.20$
Biserial Correlation Coefficients (r_{bis}):	$r_{bis} \geq 0.20$

Additional criteria, which were the reading skills as specified in the test specifications, were also applied in selecting the test items in the reading comprehension section. In addition to those criteria mentioned, the findings from the pilot questionnaire and related studies concerning health problems such as eyestrain, fatigue associated with staring at a screen (Mourant et al, 1981 cited in Clariana and Wallace, 2002), and cognitive workload that may be caused by the use of a new test task type (Noyes et al, 2004) were also applied in the consideration.

The selected items were those that obtained the aforementioned statistic values required. For the reading comprehension section, the selected test items were those that produced both the required statistic values and those that were matched in terms of the skill being tested. Some test items in this section of the two modes that were not eligible statistically but were matched based on the measured skills were also selected. Those items were then revised and used in the main data collection procedure.

Based on those criteria, there were 20 vocabulary items selected. Twelve vocabulary items were selected from the paper-based mode and eight items were from the computer-based one. To obtain ten items in each mode, two items from the paper-pencil mode were used in the computer-based one. Among the 60 reading comprehension test items, the test items from each passage that were found to be parallel were those that aimed to measure six reading skills, namely scanning (finding specific information), finding pronoun references, guessing word meaning from context, finding main ideas, making inferences, and understanding explicitly and implicitly stated ideas. Only 24 items in the reading comprehension section were parallel. This means that there were six parallel test items corresponding to each

reading passage. Thus, the total number of the selected test items administered in the main study was 44 out of 100. The statistic values of the selected items are shown in Appendix D.

Of 44 test items selected, 22 items were administered conventionally and the other 22 items were delivered through the computer. Of the 22 test items in each mode, ten items were in the vocabulary section and twelve items were in the reading comprehension part. The data obtained from the main study was then analyzed using the CTIA/Grading program.

2.2.4 The quantitative analysis of the questionnaire

Each participant was asked to respond to the questionnaire before and after doing the tests in the two modes. The questionnaire is in the five-point Likert's scale and the total number of the items is 35. The pilot questionnaire is shown in Appendix E. The scale of one refers to strong disagreement or the most negative attitude while the scale of five refers to strong agreement or the most positive attitude.

The questionnaire is composed of four parts: Part 1 on the students' anxiety towards English language tests, Part 2 on the students' familiarity with the computer, Part 3 on the students' attitudes towards the paper-based and the computer-based of the reading test of English, and Part 4 on the students' preferences towards modes of the reading test of English, which is in an open-ended format. All four parts were distributed to the participants before they did the tests. The set that was distributed before the test administration is called the pre-test questionnaire. Only Parts 3 and 4 were distributed after they completed the tests from the two modes, and this set is named the post-test questionnaire. The data from the pre-test and post-test were analyzed separately.

Although the number of the students who completed the RCTE from the two modes is 35, only 34 sets of the questionnaire were answered prior to the test administration and 33 sets of the questionnaire were collected subsequent to the test administration. The responses from the questionnaire were analyzed using SPSS for Windows to find descriptive statistics, frequency, reliability, and correlations. The statistics values are illustrated in Table 7 on page 84. The grand mean score of all 35 items is 117.03 with standard deviation of 9.17 (N=33). The alpha reliability coefficient of the entire questionnaire is .66.

Part 1: English language test anxiety. This section is composed of nine items. The mean score of this section is 29.61 with a standard deviation of 2.47. The reliability alpha coefficient is achieved at -.26.

Part 2: Familiarity with the computer. This section is also composed of nine items. The mean score of this section is 30.64 with a standard deviation of 5.23. The alpha reliability coefficient is .75.

Part 3: Attitudes towards the paper-based and the computer-based RCTE. This section is composed of seventeen items and was distributed before and after the participants did the tests in the two modes. The mean score from the pre-test is 56.76 with a standard deviation of 4.76 (N=34), while the post-test yields a mean score of 56.24 with a standard deviation of 5.75 (N=33). The alpha reliability coefficient of the pre-test is .41 and the alpha reliability coefficient of the post-test is .51.

Table 7. Mean Scores and Reliability Coefficients of the Pilot Questionnaire

Part	\bar{X}	Standard Deviation (SD)	K-R20	Correspondents
All 35 items	117.03	9.17	.66	33
Part 1 (9 items)	29.61	2.47	-.26	34
Part 2 (9 items)	30.64	5.23	.75	34
Part 3 (17 items)				
Part 3 (Pre)	56.76	4.76	.41	34
Part 3 (Post)	56.24	5.75	.51	33

From the statistic values illustrated in Table 7 above, it is apparent that Part 1 of the questionnaire produced a very low and negative internal consistency estimate (K-R20 = -.26). Although this part was to find if the students had anxiety towards English language tests in general, the students might be confused with the RCTE tests that they had to take. To avoid this ambiguity, which could contaminate the study results, this part was discarded.

Part 3 of the questionnaire was revised based on the same constructs. Two additional lecturers from the English Department at CMU were asked to review the revised questionnaire. The revised version consists of ten main topics. The total number of items is 26. Of 26 items, eleven items are related to the paper-and-pencil test and twelve items are about the computer-based test. Three other items are about

the students' familiarity with other computerized tests and the language of the test instructions. The revised questionnaire is shown in Appendix G. The alpha reliability coefficient of all 26 items is .69 in the pre-test and .73 in the post-test. These reliability coefficient values are obtained from the main study.

Therefore, the questionnaire used in collecting the data for testing the second hypothesis is composed of three parts: Part 1 on the students' familiarity with the computer, Part 2 on the students' attitudes towards the conventional and the computer test modes, and Part 3 on their general opinions. The total number of the items is 35.

In sum, the RCTE of 100 test items and the attitude questionnaire underwent both qualitative and quantitative analyses. The reading passages used in the reading comprehension section have readability levels ranging from grade levels 10.7 to 12 and their reading ease levels ranging from 39.1 to 53.2. The reading passages can be considered as difficult based on Flesch's scales. The 100 items were split into halves. That is 50 items were used in the paper-pencil test form and the other 50 items were computerized. The paper-pencil set has an internal consistency estimate of .87 and the computerized set has an internal consistency estimate of .83. To make the two sets equivalent and comparable, the test items were examined for similar statistic values. In addition, those items in the reading comprehension section were examined for similar reading skills. Therefore, only 44 test items were selected and used to collect the data. The reading skills were also matched for similar sub skills, which are scanning, finding pronoun references, guessing word meaning from context, finding main ideas, making inferences, and understanding explicitly and implicitly stated ideas.

Regarding the attitude questionnaire, Part 1 produced a negative internal consistency estimate of $-.26$ while Part 2 yielded a moderate internal consistency estimate of $.75$. Part 3, which compared the students' attitudes towards the two modes of testing before and after the test administrations, produced low reliability coefficients. Before the test administration, Part 3 had an internal consistency estimate of $.41$. After the test administration, it produced an internal consistency estimate of $.51$. As a result of these low internal consistency estimates Part 3 was then revised. The revised attitude questionnaire was composed of three parts: Part 1 on students' familiarity with computers, Part 2 on students' attitudes towards the two modes of testing (the paper-pencil and the computer-based tests), and Part 3 on students' general opinions.

The data collection and data analyses

1. Data collection procedure and design

The paper-and-pencil RCTE and the computer-based RCTE were administered to 66 students. The students were randomly divided using the counterbalanced design (see Table 8 below). This means they sat for both tests in the two modes of testing. The answers that the students made in the paper-and-pencil test were compiled and recorded on a prepared item analysis sheet and the responses from the computer-based test were recorded on diskettes. The computer-based RCTE raw data was then printed out and recorded on the prepared sheet for the item analysis. The scale scores from the questionnaire were recorded in the Statistical Package for the Social Sciences Software (SPSS) for Window version 11.5 to be analyzed.

Table 8. Design for Test Administrations

Pre-test Questionnaire	Paper-and-pencil RCTE	Computer-based RCTE	Post-test Questionnaire
	Group A	Group B	
	Group B	Group A	

2. Data analyses

The raw scores from both the conventional RCTE and the computer-based RCTE were analyzed statistically using the SPSS program for Window version 11.5. The total scores from each test mode were analyzed for descriptive statistics. The correlation coefficient between the two sets of tests was also calculated using Pearson Product Moment Correlation. The correlation coefficient between the RCTE of the two modes and the English test scores from the Entrance Exam was set at .70 (relatively moderate relationship (Brown, 1996: 153)). The alpha or significance level was set at .05. To test the hypotheses, t-test (dependent sample) was applied.

2.1 The quantitative analysis of the RCTE

The total number of test items administered conventionally and by computers was 44 items (i.e. 22 items in the paper-and-pencil form and 22 items in the computer-based form). The grand mean of all 44 items is 19.29 with the standard deviation of 7.16 and the internal consistency estimate (K-R20) is .83 (N=66). When analyzing the data from each test mode separately, the mean score of the paper-based version is 9.7 with a standard deviation of 4.0 and it obtains an internal consistency estimate

(K-R20) of .74. Similarly, the computer-based set produces the low mean score of 9.5 with the standard deviation of 4.0 and its K-R20 reliability coefficient is .72 as shown in Table 9.

Table 9. Descriptive Statistics of the RCTE

Mode of presentation	Number of items	Total scores	\bar{X}	SD	K-R20
Both (P&P and Computer)	44	44	19.29	7.16	.83
P&P	22	22	9.7	4.0	.74
Computer	22	22	9.5	4.0	.72

N = 66

Note: These descriptive statistics are obtained from the item analysis using the CTIA.

The scores from each part from the two modes were calculated for correlations using the Pearson Product Moment Correlation Coefficient formula. Table 10 on page 89 shows that the paper-based vocabulary section obtains a moderate and significant correlation with the reading comprehension section and with the computer-based vocabulary at .523 and .515 respectively. Similarly, the computer-based vocabulary section is significantly correlated with the paper-based reading comprehension section and the computer-based reading comprehension section at .516 and .575 respectively. The paper-based vocabulary section reveals a significant correlation with the computer-based reading comprehension section at a low level. The correlation coefficient is .371. The reading comprehension sections of both modes also yield a low but significant correlation at .442.

In addition, each subsection produces significant correlations with the total score of each mode. The four subsections yield from a moderate and significant correlation to a high and significant correlation with the total score of the conventional mode. The correlation coefficients range from .470 to .902. Similar correlations are also found between the four subsections and the total score of the computer-based mode. The correlation coefficients for this mode range from .494 to .901. It is also apparent that the two subsections of each mode produce a high and significant correlation with their respective mode of test presentation. However, the four subsections reveal a generally moderate and significant correlation with the scores from the Entrance Examination. The correlation coefficients range from .533 to .692.

Table 10. Correlation Matrix of the Paper-Based RCTE, the Computer-Based RCTE, and the Entrance Examination (EE) scores

	P&P VC	P&P RC	CB-VC	CB-RC	P&P RCTE	CB- RCTE	EE scores
P&P VC	--	.523**	.515**	.371**	.839**	.494**	.533**
P&P RC		--	.516**	.442**	.902**	.537**	.596**
CB-VC			--	.575**	.590**	.873**	.692**
CB-RC				--	.470**	.901**	.645**
P&P RCTE					--	.593**	.660**
CB- RCTE						--	.752**
EE scores							--

N = 66, N of items = 44, **. $p \leq .01$ (two-tailed).

The total scores gained from the two modes were also calculated to find if there were any relationships with each other and with the students' scores of the

English language test from the Entrance Examination. The result shows that the paper-based RCTE is moderately and significantly correlated with the computer-based RCTE at .59. When examining the relationship of each mode with the English test scores from the Entrance Exam, the paper-based RCTE produces a moderately high and significant correlation with the English test scores from the Entrance Exam ($r = .660$). Interestingly, the computer-based RCTE yields a higher significant correlation with the English test scores from the Entrance Exam than the conventional mode at .75. This is probably because the students were motivated in doing both tests.

2.2 The quantitative analysis of the questionnaire

The questionnaire that was distributed in the main study consisted of three parts. Part 1 was distributed to the students before the test administration in order to survey their familiarity with the computer. Parts 2 and 3 on the students' attitudes towards the two modes of testing were distributed to the participants before and after they took the tests from the two modes. The statistic values of Part 2 from before and after the test administration were analyzed separately and in combination. Table 11 on page 91 shows that the mean score of all 35 items is 114.68 with a standard deviation of 13.29. The entire set yields an internal consistency estimate of .80.

Part 1: The students' familiarity with the computer. The mean score of this part (9 items) is 29.60 with a standard deviation of 6.5. The K-R20 reliability coefficient of this part is high at .87. The results show that the students were quite familiar with the computer and its use.

Part 2: The students' attitudes towards the RCTE from the two modes of testing. This part consists of 26 items and is divided into two subparts: attitudes about

the paper-based test (11 items) and attitudes about the computer-based test (12 items). The other three items are concerned with other features of the test (i.e. the students' familiarity with any computer-based tests of English and the language of the instructions).

The pre-test set on the paper-based test yields a mean score of 36.47 with a standard deviation of 4.9 and the K-R20 reliability coefficient of those 11 items is moderate at .55. The post-test result of the same set produces a slightly higher mean score of 36.83 with a standard deviation of 5.8, and the K-R20 reliability coefficient increases to moderately high at .73.

Table 11. Descriptive Statistics and Reliability Coefficient of the Questionnaire

	\bar{X}	SD	K-R 20
Parts 1 & 2 (Before test administrations; 35 items)	114.68	13.29	.80
Part 1 (Before test administrations; 9 items)	29.60	6.5	.87
Part 2 (Before test administrations; all 26 items)	85.07	9.34	.69
a) Attitudes towards P&P test before test administrations	36.47	4.9	.55
b) Attitudes towards computer-based test before test administrations	39.98	5.2	.58
Part 2 (After test administrations; all 26 items)	87.74	9.93	.73
a) Attitudes towards P&P test after test administrations	36.83	5.8	.73
b) Attitudes towards computer-based test after test administrations	42.04	5.6	.65

N = 66

Regarding the attitudes towards the computer-based test, the pre-test result reveals a high mean score of 39.98 with a standard deviation of 5.2, and the K-R20

reliability coefficient is moderate at .58. The post-test result of the same questionnaire set produces a higher mean score of 42.04 with a standard deviation of 5.6 and the K-R20 reliability coefficient obtained is moderate at .65.

It is apparent that the mean score of the post-test on the attitudes towards the computer-based test is higher than that of the pre-test. In other words, the students appeared to change their perceptions and preferences about the computer-based test once they completed the actual computer-based test. Apparently, the mean scores of the attitudes towards the computerized test that were gathered before and after the administration of the test are higher than the attitudes towards the paper-and-pencil test.

In sum, the item analyses reveal that the RCTE of 44 items, regardless of mode of test presentation, yield an internal consistency estimate of .83. The paper-and-pencil test has an internal consistency estimate of .74 and the computer-based test has an internal consistency estimate of .72. Pearson Product Moment Correlation was carried out to find the correlations between the paper-pencil RCTE, the computer-based RCTE, and the English scores from the EE. The results of this analysis show that the paper-pencil RCTE has a significant but low correlation with the computer-based RCTE at .593 level. The paper-pencil RCTE also yields a significant but low correlation with the English scores from the EE at the level of .660. However, the computer-based RCTE yields a significant and moderate correlation with the English scores from the EE at the level of .752.

The attitude questionnaire, which was composed of three parts with 35 items, also underwent statistical analyses. The three parts were analyzed in combination and separately. The whole set produces a high internal consistency estimate at .80. Part

1, which consisted of nine items on the students' familiarity with computers, also yields a high internal consistency estimate at .87. Part 2, which was divided into two subsections and consisted of 26 items, was distributed before and after the test administrations. Therefore, the raw data of this part was analyzed separately. Before the test administration, Part 2 yields a low internal consistency estimate of .55 while it yields a moderate internal consistency estimate of .73 after the administration.

Summary

One of the important issues of this study is to validate the instruments, namely the RCTE which was delivered through two modes of testing (the conventional and the computer) and the attitude questionnaire. The research design, population and sample of the study were illustrated in detail together with the procedures of the data collection and data analysis. To validate the RCTE, validation procedures were applied in its development stage to the procedures of collecting empirical data from the pilot and the main studies.

More importantly, the study aims to find out whether or not the computer mode of assessment affects reading comprehension test performances. The t-test was used to compare the mean scores from the conventional and the computer-based RCTE. The other aim of the study is to find out students' reactions to the two modes of testing. The mean scores from the questionnaire, which was distributed before and after the students took the tests in the two modes, were also compared. The results of the data analyses and the answers to the two hypotheses are presented in the following chapter.

CHAPTER IV

FINDINGS

Chapter IV briefly presents a summary of the instrument validation and results from the study. The findings from the hypothesis testing are demonstrated and findings from the questionnaire are also presented. The last part deals with the summary and discussion of the findings.

Summary of the instrument validation

The instruments, namely the RCTE and the attitude questionnaire, were examined for reliability and validity as demonstrated in the previous chapter. The RCTE was split into two halves: the paper-and-pencil form and the computer form. The two forms were found to be equivalent in terms of their statistical values and reading skills. The paper-and-pencil test form has an internal consistency estimate of .74 and the computer-based test form has an internal consistency estimate of .72. The attitude questionnaire of 35 items obtained a high internal consistency estimate of .80. These two instruments were then implemented in collecting the data. The following section presents the results of hypothesis testing.

Hypothesis testing

The research questions of the study stated in the first Chapter are (1) to what extent the computer-based reading test of English measures the reading ability of first year university students and (2) what the students' attitudes towards the conventional

and the computer-based reading tests of English are. Based on the literature review in Chapter II, the first hypothesis was stated in the null form while the second hypothesis was posited in the alternative form as follows:

H_1 : There is no significant difference between the mean score of the computer-based reading comprehension test of English and that of the conventional reading comprehension test of English.

H_2 : The mean score of the attitudes towards the computer-based test is higher than that of the conventional test.

1. Hypothesis 1 testing

The first hypothesis is that there is no significant difference between the mean score of the computer-based RCTE and that of the conventional RCTE. In other words, the mean score from the computer-based test is the same as that from the conventional test ($\bar{X}_{\text{CBT}} = \bar{X}_{\text{P\&P}}$). In order to test this hypothesis, the mean scores from the paper-based and the computer-based tests were compared using the t-test of dependent sample formula. The confidence interval of the difference (p value) was set at .05 (95%). The results show that the t value is -.57 with a degree of freedom of 65. The critical t value for 65 is 1.994 ($\alpha = .05$). Hence, it is apparent that the mean score of the computer-based test does not exceed the critical value. Therefore, there is no significant difference between the mean score from the computer-based test and that from the conventional test. It can be said that the computer-based test can measure reading ability as well as the conventional test. The result of the t-test is shown in Table 12 below.

Table 12. A Comparison of the Mean Scores from the Paper-Based RCTE and the Computer-Based RCTE

Mode of assessment	\bar{X}	Mean Differences	t	df	Sig. (2-tailed)
CB-RCTE	9.42	-.26	-.57	65	.57
P&P RCTE	9.68				

$\alpha < .05$, $p < .05$, $N = 66$

2. Hypothesis 2 testing

The second hypothesis purported that the mean score of the attitudes towards the computer-based test is higher than that of the conventional test. ($\bar{X}_{\text{attitudes to the CBT}} > \bar{X}_{\text{attitudes to the P\&P}}$). To test the hypothesis, the t-test of dependent sample was applied. The subjects were treated as a self-controlled group in this study.

The scores of Part 2 of the questionnaire that were collected before and after the test administrations were analyzed separately and compared. Moreover, within this part there were two subsets of items: attitudes towards the paper-pencil test and those towards the computerized test. Therefore, there were four sets of mean scores to be compared (see Table 13 on page 98).

Pair one: the mean scores from the pre-test questionnaire on the conventional and the computer modes were compared. The result of the t-test shows that the mean difference is -3.51 . The t value with the degree of freedom of 65 is -5.05 . This shows that the students revealed more positive attitudes towards the computer-based test before they took the tests on both modes.

Pair two: the mean scores from the post-test questionnaire on the conventional and the computer modes were compared. The result of the t-test shows that the mean

difference is -5.21 and the t value with the degree of freedom of 65 is obtained at -6.22 . After the test administrations, it is apparent that the students had positive attitudes towards the computer-based test.

Pair three: the mean scores of the attitudes towards the paper-pencil mode before and after the test administrations were compared. The t -test result of the attitudes towards the paper-pencil mode does not exceed the given value for the .05 level (t value = $-.66$). This means that the students did not alter their attitudes towards the paper-pencil mode before and after they completed the tests.

Pair four: the mean scores of the attitudes towards the computer mode before and after the test administrations were compared. The t -test result of the attitudes towards the computer mode does exceed the given value of the .05 level. It means that the students' attitude towards the computer mode improved over the course of the test administration. This may be because the computer mode attracts the students' attention due to its unconventional format and its interactiveness.

It is apparent that the t values of the two t -test results exceed the critical value ($df = 65$, critical t value = 1.99). This means that there is a significant difference in the students' attitudes towards the conventional and the computer modes of testing before and after they completed the tests. In other words, the students showed more positive attitudes towards the computer-based test than towards the conventional test before and after they took the tests.

Table 13 (p. 98) shows the comparison of mean scores of Part 2 of the questionnaire that was collected before and after the test administrations.

Table 13. A Comparison of the Mean Scores of Part 2 of the Questionnaire
Before and After the Test Administrations

	\bar{X}	Mean Difference	t	df	Sig. (2-tailed)
Pre-test on P&P	36.47				
Pre-test on CBT	39.98	-3.51	-5.05	65	.000
Post-test on P&P	36.83				
Post-test on CBT	42.04	-5.21	-6.22	65	.000
Pre-test on P&P	36.47				
Post-test on P&P	36.83	-.36	-.66	65	.506
Pre-test on CBT	39.98				
Post-test on CBT	42.04	-2.06	5.76	65	.005

$\alpha < .05$, $p < .05$, $N = 66$

Note: All students took both tests. In this study they were used as both controlled and experimental group.

Findings from the questionnaire

The attitude questionnaire is in the five-point Likert's scale and there are five possible degrees that may be chosen. The interval length was calculated so that the questionnaire can be interpreted comprehensively and constantly. The following formula of Best and Kahn (1993) was applied.

$$\frac{\text{The highest score} - \text{the lowest score}}{\text{The number of intervals}} = \frac{5 - 1}{5} = 0.8$$

The range of .8 between each interval is used to interpret the mean score as shown in Table 14 on page 99. The mean score of each questionnaire item is interpreted according to the label value.

Table 14. The Interpretation of the Questionnaire

\bar{X}	Ability level /Frequency (Part 1)	Degree of agreement (Part 2)
1.00—1.80	Very poor/Not at all	Strongly disagree (SD)
1.81—2.60	Poor/Rarely	Disagree (D)
2.61—3.40	Average/Sometimes	Undecided (U)
3.41—4.20	Good/Frequent	Agree (A)
4.21—5.00	Excellent /Always	Strongly agree (SA)

1. The interpretation of Part 1: students' familiarity with the computer

From the mean scores shown in Appendix H, it can be said that the students in the study had good opportunities to access computers. In other words, they appeared to be average computer users who got access to the computer regularly ($\bar{X} = 3.28 - 3.30$). The places that they often got access to the computer were either at the computer labs provided on campus or at home. They also had high capability in using basic computer devices and commands such as using a Thai keyboard to type Thai ($\bar{X} = 3.50$), using a mouse or a touch pad ($\bar{X} = 3.69$), and searching the Internet ($\bar{X} = 3.89 - 4.10$). However, their ability in using some programs to create graphic designs, pictures, and presentations was quite low ($\bar{X} = 2.48$).

2. The interpretation of Part 2: students' attitudes towards the conventional and computerized RCTE

The interpretation of this part is based on the mean scores illustrated in Appendix H. The responses relating to familiarity with any computerized tests indicated that the students were not familiar with this kind of test before and after the

RCTE administrations ($\bar{X}_{\text{before}} = 2.15$, $\bar{X}_{\text{after}} = 2.15$). When asked whether they had any concerns about the test presented on paper and computers, the students revealed that they were quite worried about taking the test regardless of the mode of test presentation. Before taking the tests, they appeared to be more concerned about the computerized test than the paper-pencil one ($\bar{X}_{\text{P\&P before}} = 3.06$, $\bar{X}_{\text{CBT before}} = 3.12$). After taking the tests, their concerns increased ($\bar{X}_{\text{P\&P after}} = 3.18$, $\bar{X}_{\text{CBT after}} = 3.18$). The students became more concerned about the paper-and-pencil test after they finished the test. However, their concerns about the computerized test increased only slightly.

The students also altered their opinions about the format of the test items. Before the test administration, they showed a high preference for the multiple-choice (MC) format, which was used in both modes ($\bar{X}_{\text{P\&P before}} = 3.63$, $\bar{X}_{\text{CBT before}} = 3.54$). However, after the test administration, their preference towards the MC format in both modes dropped ($\bar{X}_{\text{P\&P after}} = 3.54$, $\bar{X}_{\text{CBT after}} = 3.46$). The student's attitudes towards the length of the reading passages seemed to vary based on the mode of the test presentation. The students showed positive attitudes towards the reading passages of about 350-400 words if they appeared on paper. Their positive preference on this was not altered before and after the test administrations ($\bar{X}_{\text{P\&P before}} = 3.19$, $\bar{X}_{\text{P\&P after}} = 3.15$). Despite the fact that the reading passages in both modes were about the same length, the students still felt that such length was not appropriate for them to read on a screen ($\bar{X}_{\text{CBT before}} = 2.43$, $\bar{X}_{\text{CBT after}} = 2.87$). The students agreed that the font size of Times New Roman 13 was suitable for use in both tests. However, they showed more agreement with the use of that font in the computerized test ($\bar{X}_{\text{CBT before}} = 3.45$,

$\bar{X}_{\text{CBT after}} = 3.53$). When checking the students' preferences on the language of the instructions, the students showed mainly positive attitudes towards the Thai instructions both before and after they took the tests.

It is interesting that the students perceived the paper-and-pencil and the computerized tests as difficult before they took the tests ($\bar{X}_{\text{P\&P before}} = 3.37$, $\bar{X}_{\text{CBT before}} = 3.54$) but after the test administrations they thought that the paper-and-pencil test was more difficult than the computerized test ($\bar{X}_{\text{P\&P after}} = 3.68$, $\bar{X}_{\text{CBT after}} = 3.54$). This implies that the students might not feel very stressed when doing the computerized test. Moreover, the students revealed that the time allotment of one hour was not sufficient for them to complete the paper-and-pencil test ($\bar{X}_{\text{P\&P before}} = 2.77$, $\bar{X}_{\text{P\&P after}} = 2.92$). However, they agreed that they could finish the computerized test in one hour as allowed ($\bar{X}_{\text{CBT before}} = 3.04$, $\bar{X}_{\text{CBT after}} = 3.04$).

When examining the students' preference for the distinctive features of the conventional mode, before and after taking the test, the students tended to prefer to see all of the test items at once ($\bar{X}_{\text{P\&P before}} = 3.78$ and $\bar{X}_{\text{P\&P after}} = 3.63$) and they also preferred to take notes ($\bar{X}_{\text{P\&P before}} = 4.07$ and $\bar{X}_{\text{P\&P after}} = 3.98$). The features of the computerized test that the students tended to prefer are moving pages by clicking navigating icons ($\bar{X}_{\text{CBT before}} = 3.43$ and $\bar{X}_{\text{CBT after}} = 3.78$) and precise marking of the computer program ($\bar{X}_{\text{CBT before}} = 3.57$ and $\bar{X}_{\text{CBT after}} = 3.80$). However, they showed neutral attitudes towards some features of the paper-pencil test mode. These were turning pages manually ($\bar{X}_{\text{P\&P before}} = 3.22$ and $\bar{X}_{\text{P\&P after}} = 3.25$), viewing the list of items ($\bar{X}_{\text{P\&P before}} = 3.34$ and $\bar{X}_{\text{P\&P after}} = 3.36$), and marking by teachers instead of by

the computer ($\bar{X}_{\text{P\&P before}} = 2.86$ and $\bar{X}_{\text{P\&P after}} = 2.90$). In addition, they seemed uncertain whether they felt active while doing the computer-based test ($\bar{X}_{\text{CBT before}} = 3.06$ and $\bar{X}_{\text{CBT after}} = 3.31$).

Nevertheless, the students did alter their preferences about the computer-based test after they completed it. They showed preferences for moving by scrolling ($\bar{X}_{\text{CBT before}} = 3.34$ and $\bar{X}_{\text{CBT after}} = 3.63$), viewing each reading passage with one test item on the other side ($\bar{X}_{\text{CBT before}} = 3.22$ and $\bar{X}_{\text{CBT after}} = 3.65$), and receiving an immediate score report ($\bar{X}_{\text{CBT before}} = 3.18$ and $\bar{X}_{\text{CBT after}} = 4.19$).

3. Findings from Part 3 of the questionnaire

This part allowed the students to freely express their opinions about the test on the two modes. Their statements can be summarized as follows:

The paper-pencil test. They reported that they were familiar with this mode of testing and some of them felt that they could control the test because they did not have to worry about technical problems which might occur with the computer-based test. They also emphasized that when doing the paper-pencil test they were able to take notes and highlight key words or important statements. They reported that they could start at any test item they wanted to. In other words, they did not have to proceed with the test items in sequence. Interestingly, they seemed to use more test-taking strategies in doing the paper-pencil test.

However, they mentioned that they did not like the paper-pencil test due to the number of pages of the test booklet. They stated that they did not like turning pages because it could confuse them, especially when the reading passages and test items were on separate pages.

The computer-based test. The respondents stated that the computer-based test was convenient for them and they did not have problems with turning pages as in the paper-based test. They liked the computer-based test because it reported their scores immediately after they finished the test. Some reported that the computer-based test and its features such as clicking and movements helped them stay alert and active while doing the test.

Nevertheless, they all agreed that staring at the screen to read and answer the test items irritated their eyes. They also pointed out certain technical problems that might occur during the test session, such as sudden blackout and problems with the computer hard drives and storage processes.

Summary

The findings of the study, namely the test performance of the computerized and conventional RCTE and the responses from the attitudes questionnaire, were statistically analyzed and used to test the hypotheses. The hypothesis testing regarding the result of the RCTE revealed that the test performance from the computerized RCTE was not significantly different from that of the conventional RCTE. Another hypothesis testing regarding the attitudes towards the RCTE delivered through the computer and conventional modes also revealed that the participants showed more positive attitudes towards the computerized RCTE than the conventional RCTE before and after the test administrations.

CHAPTER V

DISCUSSION AND RECOMMENDATIONS

Chapter V presents a summary and the results of the study. It is then followed by discussion of the findings before the conclusions are presented. The chapter ends with recommendations concerning the implementation of computerized tests in the educational context and for future studies.

Summary of the study

1. Purposes of the study

This study was initiated to investigate the effects of test presentation modes on English reading comprehension ability and test performance of students who were studying English as a foreign language. Specifically, the purposes of the study were (1) to compare the performances of the first year university students on the English reading comprehension test in the conventional and computer modes of testing and (2) to compare the students' attitudes towards the two modes of testing.

The underlying issues of the study were to validate and investigate the equivalence of the two sets of the Reading Comprehension Test of English (RCTE) and to find out if reading comprehension test performances are affected by the use of computers in testing.

2. Participants of the study

The participants in this study were a group of 66 first year students from the Faculty of Humanities at Chiang Mai University (CMU). They were purposefully selected from intact groups. The English language proficiency level of 35 of them

was in the intermediate level, 23 of them were in the low level, and only eight were in the high level of English proficiency. They were majoring in French, Mass Communication, Home and Community, English, Thai, Philosophy, Japanese, and Chinese.

3. The instruments

The instruments were the RCTE and the attitude questionnaire. The RCTE was developed and delivered through two different modes of testing (the paper-and-pencil and the computer modes). The delivery method of the computer mode was not adaptive. It was composed of two separate test sets. The two sets were written according to the same test specifications. The paper-and-pencil RCTE had the internal consistency estimate of .74 and the computer-based RCTE had the internal consistency estimate of .72. The attitude questionnaire was of one set and it produced the internal consistency estimate of .80. The questionnaire was composed of three parts. Part 1 asked about the students' familiarity with the computer. Parts 2 and 3, which elicited the students' attitudes towards the two modes of testing, were distributed before and after the test administration to find out whether the students altered their attitudes before and after taking the two forms of RCTE.

4. Procedures of the study

The study was conducted in two phases. In phase one, the instruments were developed and underwent qualitative validation procedures during October 2003 – February 2004 at Chulalongkorn University. In phase two, the pilot and the main studies were carried out at CMU during June-August 2004. The pilot study was performed to validate the instruments quantitatively. The instruments were then

revised. The main study was conducted in September 2004 to collect data and test the hypotheses.

5. Data collection and analyses

The data collected from the two studies were analyzed at Chulalongkorn University. The RCTE items underwent item analyses using the CTIA program at the Academic Testing Center of Chulalongkorn University in order to validate the test quantitatively. The scores of each test mode were analyzed both separately and in combination. Moreover, Pearson Product Moment Correlation was performed to find the relationships between the paper-and-pencil RCTE, the computerized RCTE, and the English test scores from the Entrance Exam. The data from the attitude questionnaire was analyzed using the SPSS program for descriptive statistics, correlation coefficients, and alpha reliability coefficients. The t-test (dependent) sample was applied to compare the mean of the paper-and-pencil test and that of the computerized test and to compare the means of attitude scores from the questionnaire administered before and after taking the tests.

Summary of findings

The findings of the study are in two major areas: (1) the test performances from the two modes of testing and (2) the attitudes towards the two modes of testing.

1. Findings on the test performances

The first hypothesis regarded the equivalence of the mean scores from the tests delivered through the computer and the paper-and-pencil modes. The t-test formula of dependent sample was applied and the result revealed that there was no significant difference between the mean score produced from the computer-based test and that

from the conventional test. Therefore, the null hypothesis, which stated that there was no significant difference between the mean score of the computer-based test and the mean score of the conventional test was accepted.

2. Findings on the attitudes towards the two modes of testing

The second hypothesis concerned the students' attitudes towards the two modes of testing. It stated that the mean score of the attitudes towards the computer-based test was higher than that towards the conventional test. The t-test formula of dependent sample was performed. The mean scores of the questionnaire distributed before and after the test administrations were compared.

The mean scores of the attitudes of the students towards the two modes of testing prior to the test administration were firstly compared. The t-test result revealed that there was a significant difference between the mean score of the attitudes towards the computer-based mode and that towards the conventional mode before they took the tests.

Apart from that, the mean scores of the attitudes towards the two modes obtained subsequently to the test administrations were compared. The result also indicated that there was a significant difference between the mean score of the attitude towards the computer-based mode and that towards the conventional mode after the students completed the tests.

When examining the mean scores closely, it is apparent that the mean scores of the attitudes towards the computer-based mode obtained before and after the test administrations were higher than those towards the conventional mode. It can be inferred that the students who participated in the study had positive attitudes towards

the computer-based test. In other words, the students showed a greater preference for the computer-based test than for the conventional one.

This implies that the computer-based test can measure the reading ability as well as the conventional test. Also, in this study the students preferred the computer-based test to the conventional test before and after taking the tests.

Discussion of the findings

The purposes of the study were (1) to compare the test performances from the conventional and the computer-based tests and (2) to compare the students' attitudes towards the two modes of testing. Hence, the two hypotheses were set. The first hypothesis stated that there was no significant difference between the mean score from the computer-based test and that from the conventional test. The second hypothesis was set stating that the mean score of the attitudes towards the computer-based test was higher than that of the conventional test. The test performances were obtained from the RCTE specifically designed for this study and the attitude results were obtained from the attitude questionnaire. The findings of the study are discussed according to the two aforementioned hypotheses as follows.

1. The RCTE results

1.1 Comparison of the test scores from the computerized and conventional RCTE

The results of the t-test of dependent sample formula revealed that the test performances of the students on the two modes of testing were not statistically different ($t = .57$, $df = 65$, $p > .05$). This means that the students did not perform better or worse on the computer-based test than they did on the paper-pencil test

($\bar{X}_{\text{computer}} = 9.42$, $\bar{X}_{\text{P\&P}} = 9.68$). Therefore, the first hypothesis, which stated that there was no significant difference between the mean score from the computer-based test and that from the conventional test, was accepted. This result is consistent with the studies conducted by Zandvliet and Farragher (1997), Olsen et al (1986), Muter et al (1982), Kak (1981), and Muter and Maurutto (1991) cited in Dillon (1992: 1303-1304). It can be said that the RCTE delivered through the computer can measure reading ability as well as the conventional method. From the findings, it implies that different modes of testing (i.e. computer and paper pencil modes) may not affect performance on reading tests.

1.2 The test scores from the computerized and conventional RCTE

The test performance of the students on the computer-based test is about the same as that on the conventional test ($\bar{X}_{\text{computer}} = 9.42$, $\bar{X}_{\text{P\&P}} = 9.68$). The mean scores from the two test modes appeared to be quite low and this may be due to several reasons. Referring back to the raw scores of the tests, the highest score made on the paper-pencil test was 22 out of 22 while the highest score made on the computerized test was only 18 out of 22. This difference can be a result of the students' familiarity with the conventional format. In addition, completing a reading comprehension test on a computer screen may add a greater workload to their cognitive processes (Clariana and Wallace, 2002). The students' familiarity with the computer may contribute some impact on their performances on the computerized test as was found in the studies of Belmore (1985) and Coniam (1999).

The low mean scores of the test of the two modes may also be a result of the difficulty level of the RCTE test. The students participating in the study were

considered to be mainly intermediate students and some of them were pre-intermediate students based on their English test scores from the Entrance Exam (i.e. their scores ranged from 38 to 60). Therefore, it is possible that the difficulty level of the RCTE test does not correspond with the proficiency levels of the students in the study. Another possible explanation is that they may not have been very familiar with the themes of the reading passages used in the test. Hence, their schema might not have been fully activated. Even though schema knowledge was not included in the designing stage of the RCTE, it seems that factors related to the subjects' familiarity with the content of the reading texts does have an impact on their performance (Alderson, 2000a; Barnett, 1989).

1.3 Relationships between the computerized RCTE, conventional RCTE, and the English scores from the Entrance Examination

Regarding the relationships between the paper-pencil RCTE, the computer-based RCTE, and the English test scores from the Entrance Examination, it is apparent that the paper-pencil and the computer-based RCTE have a low level of association ($r = .593$). Another unexpectedly low relationship was found between the paper-pencil RCTE and the English test scores from the Entrance Examination ($r = .660$) while the correlation coefficient between the computer-based RCTE and the Entrance Exam scores was higher at $.752$.

These are interesting results since the paper-pencil RCTE was delivered in the same manner and mode as the Entrance Examination. In addition, the computer-based RCTE appeared to have a higher relationship with the English scores from the Entrance Examination even though they were delivered in two different modes. It is possible that the reduced number of items in the main data collection caused this

decrease in the level of correlation between the paper-pencil RCTE and the English test scores from the Entrance Examination. The difficulty level of the test items of the two sets of the RCTE might have brought about this reversal. Referring back to the statistic values of the test items used in the main study, it was found that three of the vocabulary items administered on the paper test were apparently more difficult than they should have been. Therefore, this implies that the paper-pencil RCTE was slightly more difficult than the computer-based RCTE.

2. The attitudes towards the computer and conventional modes of testing

2.1 Comparison of the mean scores from the attitude questionnaire

The second hypothesis testing revealed that the students' attitudes towards the two modes of testing before and after the test administrations were significantly different (see Table 13 p. 98). The results indicated that the students had a more positive preference for the computer mode than the conventional mode before and after the administrations of the tests. The mean score of the attitudes towards the computer mode before the test administration was higher than that towards the conventional mode ($\bar{X}_{\text{computer}} = 39.98$ and $\bar{X}_{\text{P\&P}} = 36.47$). Similarly, the mean score of the attitudes towards the computer mode after the students took the test was still higher than that towards the conventional mode ($\bar{X}_{\text{computer}} = 42.04$ and $\bar{X}_{\text{P\&P}} = 36.83$). Therefore, the hypothesis stating that the mean score of the attitudes towards the computer-based test was higher than that towards the conventional test was accepted. This finding is similar to the previous studies conducted by Zandvliet and Farragher (1997) and Muter and Maurutto (1991).

2.2 Positive attitudes towards the computerized test

The subjects showed positive attitudes towards the computer presentation mode in this study. The positive preference for the use of computers in testing, particularly in this study, can be explained by two reasons. The first reason is that the students might have felt less stressed when participating in the study because they might not have associated the computerized test with the previous stressful test in the conventional mode. The other reason is that they might have been anxious about an alternative medium in testing. It cannot be concluded from this study which reason explains the positive preference for the computer mode of testing.

2.3 Additional findings from the questionnaire

Regarding the attitudes towards the two modes of testing, the students revealed an obvious preference for the computer mode. They tended to prefer only certain features of each mode. They also revealed anxiety and worry when they had to sit for English tests regardless of test presentation modes.

An additional finding from the questionnaire is that the students applied test-taking strategies when they did the tests. This can be found from the items checking on the features of the test media they preferred. They stated that they preferred seeing the questions before reading the text. On some items, they seemed not to read the passage before answering the test items. To prevent this and to measure true reading abilities, other test formats apart from the multiple-choice format should be considered in assessing reading comprehension.

In conclusion, the test performance on the computer-based RCTE was almost equivalent to that on the paper-and-pencil RCTE. In addition, the students revealed positive attitudes towards the computerized test and the use of the computer in testing.

However, there were some disadvantages such as unfamiliarity with computerized tests, program malfunctions, and eyestrain that might have affected the performance on the computerized test.

Conclusions

From the findings of this study, it can be concluded as follows:

1. The computerized RCTE and the conventional RCTE could equally assess the English reading ability of first year university students at Chiang Mai University.
2. The self-reported questionnaire results indicated that the students showed positive attitudes towards the computerized test. In other words, they preferred the computerized test to the conventional one.

Implications for language assessment

The findings from this study can be applied to language assessment in the context that English is taught as a foreign language particularly in Thailand in which English language proficiency tests are commonly required in the admission process. In higher education, computers can be used to assess students' language proficiency because they can benefit the institutions, test developers, and students. Following are some possibilities.

1. Institutions can establish their own admission tests on language proficiency and deliver the tests via computer. By doing this, the number of staff to proctor the test can be reduced and the test can be administered in one center. The computerized tests can also be integrated in regular teaching and can be used to assess

learning processes and outcomes continuously in a form of low-stakes tests or quizzes. This kind of on-going assessment provides teachers with students' strengths and weaknesses. Hence, introducing computers in the assessment procedure can promote learners' competence, ensure the quality of instruction and of the institution, as well as facilitate test administrators in administering tests.

2. Apart from the advantages to the institutions, test developers who are responsible for designing and maintaining tests can also gain a number of benefits. Working with computer programmers, test developers can create different test formats such as multiple-choice format, drag & drop, matching, true/false, and cloze. Moreover, test developers can easily revise test items and create an item bank in multiple test forms. An advanced system and high capacity in recording by computers allow testers to keep track of students' responses, time spent, and other necessary information about students' interaction with the test. Data in addition to the scores can be used to study students' test taking strategies as well as tests themselves. In short, the use of computers in testing facilitates the test developers in developing, maintaining, and studying about the tests.

3. The students themselves can also gain advantages from the computerized tests. The students in this study showed a preference for the use of computers in testing and, particularly, for the immediate score report after completing the test. Other students may also accept the computer mode of test presentation. In low-stakes test situations such as quizzes and self-assessment, students can take computerized tests at their convenience in a computer lab. Immediate score reports as well as diagnostic feedback can help students to develop their strengths and improve their weaknesses.

4. Integrating computerized tests in the classroom seems to be appealing as well. Computerized tests can be designed to be user-friendly and can enhance the regular learning process. Teachers can promote reading skills by providing gloss words, interactive features such as highlighting the meaning of content words and having visual aids together with reading passages to appear simultaneously on screen. But using the computer in assessing students' learning processes and outcomes should be done with cautions. Students should not learn to take tests just for making higher scores but they should use the test results to improve their learning.

Despite the advantages of the use of computers in testing, concerns about implementing this technology remain and may impede the implementation process. However, at present, computer technology has been rigorously improved and less complicated systems and programs are available. Therefore, the issues related to the use of the computer in assessment such as delivery systems, interface designs, and security can be fine tuned and manageable in the near future.

Recommendations for future studies

1. The designs of future studies should be experimental ones. The samples and procedures of collecting data should be strictly managed and controlled.
2. The subjects should also be assessed for their levels of proficiency by using a standardized proficiency test of the language domain aimed to measure before allowing them to participate in the study.
3. The samples' familiarity with the computer should be thoroughly evaluated in order to control the equipment familiarity effects on the test performance.

4. It will also be advantageous to provide computer training to the samples before conducting the research study.

5. In implementing computerized tests using any delivery system, it is very important that the test administrators understand the system and be able to handle technical problems that can occur during test sessions.

6. Good cooperation between technicians, computer programmers and test administrators will make the use of computers in assessment possible and successful.

7. There are also a number of issues derived from the present study that other researchers who are interested can pursue in depth. For example, there should be studies dealing with test-taking strategies on English computerized tests. Another interesting area that is not yet fully explored is reading strategies used in reading from screens in the EFL context, especially in Thailand.

8. Finally, educators should pay attention to the reading ability of Thai students and the readability levels of English language texts used in educational situations. If this area is investigated in detail, the findings will greatly assist teachers in designing reading courses that are appropriate to their students.

In conclusion, the findings of this study regarding the equivalence of the computerized and the conventional tests of reading comprehension and preferences of the students for the two modes of testing provide some insights for extensive uses of computers in both language assessment and instruction in the near future.

REFERENCES

Thai

สุพัฒน์ สุกมถันต์. 2542. การวิเคราะห์ข้อทดสอบและตัดเกรดด้วยคอมพิวเตอร์ (CTIA). กรุงเทพมหานคร: วิทยพัฒน์.

English

Aiken, Lewis R. 1998. **Tests and examinations: Measuring abilities and performance.** New York: John Wiley & Sons.

Alderson, J. C. 1984. Reading in a foreign language: a reading problem or a language problem? In J.C. Alderson and A.H. Urquhart (Eds.), **Reading in a Foreign Language**, 1-27. London: Longman.

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

Alderson, J. C. 2000b. Technology in testing: the present and future. **System** 28 (2000): 593-603.

Alderson, J. C., Clapham, Caroline, and Wall, Dianne. 1995. **Language testing construction and evaluation.** Cambridge: Cambridge University Press.

American Educational Research Association. 1999. **Standards for Educational and Psychological Testing.** Washington, DC: American Psychological Association.

Anderson, R. C., and Pearson, P. D. 1984. A schema-theoretic view of basic process in reading comprehension. In P.D. Pearson (Ed.), **Handbook of reading research**, 252-295. New York: Longman. (ERIC Document Reproduction Service. No. ED 239 236).

- Bachman, L. F. 1990. **Fundamental considerations in language testing.** Oxford: Oxford University Press.
- Bachman, L. F. 2000. Modern language testing at the turn of the century: Assuring that what we count counts. **Language Testing** 17, 1: 1-42.
- Bachman, L. F. 2003. **Validity issues in web-based language assessment.** Plenary presentation at the 5th CULI International Conference, Bangkok, Thailand, December 2003.
- Bachman, L. F. and Palmer, Adrian. 1996. **Language testing in practice.** Oxford: Oxford University Press.
- Barnett, M. A. 1989. **More than meet the eyes: Foreign language reading: Theory and practice.** Eaglewood Cliffs, New Jersey: Prentice-Hall.
- Belmore, S. 1985. Reading computer presented test. **Bulletin of the Psychonomic Society** 23: 12-14.
- Bernhardt, E. B. 1984. Three approaches to reading comprehension in intermediate German. **Modern Language Journal** 67: 111-115.
- Bernhardt, E. B. 1991. **Reading development in a second language: Theoretical, empirical and classroom perspectives.** New Jersey: Ablex Publishing Corporation.
- Bernhardt, E. B. 1999. If reading is reader-based, can there be a computer-adaptive test of reading? In Micheline Chalhoub-Deville (Ed.), **Issues in computer-adaptive testing of reading proficiency (Studies in Language Testing 10)**, 1-47. Cambridge: UCLES.
- Best, J. W. and Kahn, J. V. 1993. **Research in education.** 7th Ed. Boston: Allyn and Bacon.

- Bloom, B. S., D. E. Max, J. F. Edward, H. H. Walker, and R. K. David. 1974. **Taxonomy of educational objectives: Handbook I: Cognitive domain.** New York: David McKay Company.
- Brown, J. D. 1996. **Testing in language programs.** Eaglewood Cliffs: Prentice-Hall.
- Brown, J.D. 1997. Computers in language testing: Present research and some future directions. **Language Learning and Technology** 1, 1 (July): 44-59.
- Bugbee, A. C. Jr. 1996. The equivalence of paper-and-pencil and computer-based testing. **Journal of Research on Computing in Education** 28 (Spring): 282-299.
- Bunderson, C. V., Inouye, D. K., and Olsen, J. B. 1989. The four generations of computerized educational measurement. In R. L. Linn (Ed.), **Educational measurement**, 367-407. London: Collier Macmillan Publisher.
- Burnstein, J., Frase, L. T., Ginther, A., and Grant, L. 1996. Technologies for language assessment. **Annual Review of Applied Linguistics** 16: 240-260.
- Carrell, P. L. 1983a. Some issues in studying the role of schemata, or background knowledge, in second language comprehension. **Reading in a Foreign Language** 1, 2: 81-92.
- Carrell, P. L. 1983b. Three components of background knowledge in reading comprehension. **Language Learning** 33, 2: 183-203.
- Carrell, P. L. 1998. Some causes of text-boundedness and schema interference in ESL reading. In P. L. Carrell et al., (Eds.), **Interactive approaches to second language reading**, 101-113. Cambridge: Cambridge University Press.

- Carver, R. 1977. Toward a theory of reading comprehension and reading. **Reading Research Quarterly** 13: 8-64.
- Chalhoub-Deville, Micheline. 2001. Language testing and technology: Past and future. **Language Learning and Technology** 5, 2 (May): 95-98.
- Choi, Inn-Chull, Kim, K. S., and Boo, J. 2003. Comparability of a paper-based language test and a computer-based test. **Language Testing** 20, 3: 295-320.
- Clariana, R. and Wallace, P. 2002. Paper-based versus computer-based assessment: Key factors associated with the test mode effect. **British Journal of Educational Technology** 33, 5: 593-602.
- Coady, J. 1979. A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman, and R. R. Jordan (Eds.), **Reading in a second language: Hypotheses, organization, and practice**, 5-12. Rowley, MA: Newbury House.
- Coniam, David. 1999. Subjects' reactions to computer-based tests. **Journal of Educational Technology Systems** 27, 3: 195-206.
- Cummins, J. 1979. Linguistic interdependence and the educational development of bilingual children. **Review of educational research** 49: 222-251.
- Cummins, J. 1991. Conversational and academic language proficiency in bilingual contexts. In J. Hulstijn and A. Matter (Eds.), **AILA Review**, 8: 75-89.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., and McNamara, T. 1999. **Dictionary of language testing (Studies in Language Testing 7)**. Cambridge: Cambridge University Press.

- Davis, F. B. 1968. Research in comprehension in reading. **Reading Research Quarterly**, 3: 499-545.
- Dillon, A. 1992. Reading from paper versus screens: A critical review of the empirical literature. **Ergonomics** 35, 10: 1297-1326.
- Edelblut, P. 2003. **Examining the reliability of automated essay scoring of an ESL prompt for college writing placement.** Paper presented at the 5th CULI International Conference, Bangkok, Thailand. December 2003.
- Flesch, R. 1951. **How to test readability.** New York: Harper & Brothers.
- Fransson, A. 1984. Cramming or understanding? Effects of intrinsic and extrinsic motivation on approach to learning and test performance. In J. C. Alderson and A. H. Urquhart (Eds.), **Reading in a Foreign Language**, 86-121. London: Longman.
- Fulcher, Glenn. 2000. Computers in language testing. In P. Brett and G. Motteram (Eds.), **A Special Interest in Computers**, 93-107. Kent, UK: IATEFL.
- Gomontean, Upmiga. 1997. **Styles of Thai characters on computer screen upon the reading legibility of prothom sukka four students.** Master Degree Thesis, Audio-Visual Education, Faculty of Education, Chulalongkorn University.
- Goodman, K. S. 1967. Reading: A psycholinguistic guessing game. **Journal of the Reading Specialist** 6: 126-135.
- Goodman, K. S. (Ed.). 1968. **The psycholinguistic nature of the reading process.** Detroit: Wayne State University Press.
- Gough, P. B. 1972. One second of reading. In J. F. Kavanagh and I. G. Mattingly (Eds.), **Language by ear and by eye**, 331-358. Cambridge, MA: MIT Press.

- Grabe, William. 1988. Reassessing the term “interactive”. In P. L. Carrell et al. (Eds.), **Interactive approaches to second language reading**, 56-72. Cambridge: Cambridge University Press.
- Grabe, William. 1991. Current development in second-language reading research. **TESOL Quarterly** 25, 3: 375-406.
- Grabe, William. 2000. Reading research and its implications for reading assessment. In A. J. Kunnan (Ed.), **Fairness and validation in language assessment**. Cambridge: Cambridge University Press.
- Grabe, W., and Stoller, F. L. 2002. **Teaching and researching reading**. Harlow: Longman.
- Green, B. F. 1988. Construct validity of computer-based tests. In H. Wainer and H. I. Braun (Eds.), **Test validity**, 77-86. Hillsdale: Lawrence Erlbaum Associates.
- Just, M. A. and Carpenter, P. A. 1980. A theory of reading: From eye fixations to comprehension. **Psychological Review**, 87: 329-354.
- Kak, A. V. 1981. **Relationships between readability of printed and CRT-displayed test**. Proceedings of Human Factors Society 25th Annual Meeting (Human Factors Society, Santa Monica, CA): 137-140.
- Kenyon, D. M., and Malabonga, V. 2001. Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. **Language Learning and Technology** 5, 2: 60-83.
- Ketkham, Kanyarat. 2004. Students’ attitudes toward the computer-based reading program in the developmental reading course: The impact on their reading performances. **Thai Journal of Development Administration** 44, 2: 67-94.

- LaBerge, D., and Samuels, S. J. 1974. Toward a theory of automatic information processing in reading. **Cognitive Psychology** 6: 293-323.
- Madsen, H. S. 1991. Computer-adaptive testing of listening and reading comprehension: The Brigham Young University approach. In P. Dunkel (Ed.), **Computer-assisted language learning and testing: Research issues and practice**, 237-257. New York: Newbury House.
- Mourant, R. R., Lakshmanan, R., and Chantadisai, R. 1981. Visual fatigue and cathode ray tube display terminals. **Human Factors** 23, 5: 529-540.
- Munby, J. 1978. **Communicative syllabus design**. Cambridge: Cambridge University Press.
- Muter, P., Latremouille, S. A., Treurniet, W. C. and Beam, P. 1982. Extended reading of continuous text on television screen. **Human Factors** 24: 501-508.
- Muter, P. and Maurutto, P. 1991. Reading and skimming from computer screens and books: the paperless office revisited? **Behaviour & Information Technology** 10: 257-266.
- Noyes, J., Garland, K., and Robbins, L. 2004. Paper-based versus computer-based assessment: Is workload another test mode effect? **British Journal of Educational Technology** 35, 1: 111-113.
- Olsen, J. B., Maynes, D. D., Slawson, D., and Ho, K. 1986 (April). **Comparison and equating of paper-administered, computer-administered and computerized adaptive tests of achievement**. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

- Panitchapakdi, S. (1998). **Keynote Speech**. The 1998 International Symposium on “The Central Intellectual Property and International Trade Court as a Model for IPRs Enforcement for the 21st Century”. Bangkok, Thailand.
- Parshall, C. G., Spray, J. A., Kalohn, J., and Davey, T. 2002. **Practical considerations in computer-based testing**. New York: Springer.
- Popham, W. J. 1978. **Criterion-referenced measurement**. Eaglewood Cliffs: Prentice-Hall.
- Prapphal, K. 2003. English proficiency of Thai learners and directions of English teaching and learning in Thailand. **Journal of English Studies**, 1 (1) (2003): 6-12. Department of English, Faculty of Liberal Arts, Thammasart University, Thailand.
- Rayner, K., and Pollatsek, A. 1989. **The psychology of reading**. Eaglewood Cliffs: Prentice-Hall.
- Rumelhart, D. E. 1977. Toward an interactive model of reading. In S. Dornic (Ed.). **Attention and performance VI**. Hillsdale, NJ: Lawrence Erlbaum. 573-603.
- Rumelhart, D. E., McClelland, J. L., and the PDP Research Group. 1986. **Parallel distributed processing: Explorations in the microstructure of cognition**. 2 vols. Cambridge, MA: MIT Press.
- Sawaki, Yasuyo. 2001. Comparability of conventional and computerized tests of reading in a second language. **Language Learning and Technology** 5, 2 (May): 38-59.

- Smith, F. 1982. **Understanding reading: A psycholinguistic analysis of reading and learning** (3rd Ed.). New York: Holt, Rinehart & Winston.
- Stanovich, K. E. 1980. Toward an interactive-compensatory model of individual difference in the development of reading fluency. **Reading Research Quarterly** 16: 32-71.
- Stricker, L. J., and Wilder, G. Z. 2001. **Examinees' attitudes about the TOEFL-CBT: Possible determinants, and relationships with test performance.** (TOEFL research report 01-10. March 2001). Princeton: Educational Testing Service.
- Taylor, C., Jamieson, J., Eignor, D., and Kirsch, I. 1998. **The relationship between computer familiarity and performance on computer-based TOEFL test tasks.** TOEFL Research Report 61. Princeton: Educational Testing Service.
- Tetewornan, Paweena. 1995. **Colors and sizes of fonts on different background colors on computer screen upon reading comprehension of prathom sukka six students.** Master Degree Thesis, Audio-Visual Education, Faculty of Education, Chulalongkorn University.
- Urquhart, A. H. 1987. Comprehensions and interpretations. **Reading in a Foreign Language** 3: 387-409.
- Urquhart, A. H., and Weir, C. J. 1998. **Reading in a second language: Process, product and practice.** London: Longman.
- Wallace, C. 2001. Reading. In R. Carter and D. Nunan (Eds.). **The Cambridge guide to teaching English to speakers of other languages.** Cambridge: Cambridge University Press.

Wongboonsin, K. et al. 2003. **A study on the future of Thai workforce.** Bangkok:

National Research Council of Thailand.

Zandvliet, D., and Farragher, P. 1997. A comparison of computer-administered and

written tests. **Journal of Research on Computing in Education** 29:

423-438.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



APPENDICES

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix A

Reading Comprehension Test of the English Language (The Paper-Pencil Mode)

(ข้อสอบนี้จัดทำขึ้นเพื่อใช้ในการวิจัยระดับปริญญาโทในหัวข้อ
“การเปรียบเทียบสัมฤทธิ์ผลทางการอ่านภาษาอังกฤษด้วยวิธีการประเมินแบบคอมพิวเตอร์
และแบบประเพณีนิยม)

ข้อมูลของนักศึกษาที่เข้าร่วมในการทดสอบ			
ชื่อ		รหัสนักศึกษา	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Course <input type="checkbox"/> 103 <input type="checkbox"/> 104 <input type="checkbox"/> 203	Section		
ช่วงคะแนนสอบวิชาภาษาอังกฤษเอ็นทรานซ์หรือโควต้า	เพศ <input type="checkbox"/> หญิง <input type="checkbox"/> ชาย		
..... คะแนน			

คำชี้แจง

ข้อสอบนี้แบ่งออกเป็น 2 ส่วน ดังนี้	
ส่วนที่ 1 ข้อสอบการใช้คำศัพท์ (10 ข้อ)	หน้า 2 - 4
ส่วนที่ 2 ข้อสอบการอ่านภาษาอังกฤษ (มี 2 เนื้อเรื่อง มีคำถามเรื่องละ 6 คำถาม รวม 12 ข้อ)	หน้า 4 - 10

ขอความกรุณา

1. นักศึกษาทำข้อสอบทุกข้อ อย่างตั้งใจ
2. อย่าแยกข้อสอบออกจากกัน

หมายเหตุ

คะแนนที่ได้จากการทำข้อสอบชุดนี้จะไม่นำไปใช้ในการพิจารณาตัดเกรดในกระบวนวิชา Foundation English I, II, or III (001103, 001104, 001203)

Part I: Vocabulary (10 items; 10 points)

คำสั่ง จงอ่านข้อความแล้วเลือกคำตอบที่เหมาะสมที่ทำให้ข้อความนั้นมีความหมายสมบูรณ์ โดยวงกลมอักษร A, B, C, หรือ D หากต้องการเปลี่ยนคำตอบ ให้กากบาททับคำตอบเดิม แล้ววงกลมคำตอบใหม่ (หน้า 2-4)

1. At the end of last year, many new investors had bought shares in the stock market. They expect to receive a large sum of _____ at the end of this year for holding those shares.

- | | |
|---------------------|--------------------|
| A. salaries | B. records |
| C. dividends | D. products |

2. On Children's Day, Chiang Mai Zoo officials _____ panda dolls to children who visited the two giant pandas. The dolls were specially produced in limited number and they were not for sale.

- | | |
|---------------------|-----------------------|
| A. collected | B. illustrated |
| C. pictured | D. distributed |

3. Paradorn Srichaphan has gained himself _____ for being a polite tennis player and making rapid progress among other foreign players.

- | | |
|------------------------|----------------------|
| A. championship | B. reputation |
| C. profit | D. profession |

4. AIDS is one of many _____ diseases that has been the cause of deaths in many countries such as Swaziland. At the present, scientists and doctors still cannot find ways to prevent and stop it.

- | | |
|-------------------|----------------------|
| A. chronic | B. anonymous |
| C. novel | D. remediable |

5. Peace talks and diplomatic resolutions are first options for bringing peace to many restless or conflicting areas. However, the United Nations (UN) often has to _____ by sending multinational troops in order to protect the innocent civilians and control the fighting situations.

- | | |
|--------------------|---------------------|
| A. declare | B. consult |
| C. obstruct | D. intervene |

6. The family of Big, a member of a famous boy band called D2B, were _____ when they learned from the doctors that the chances that Big will live a normal life again were slight.

- | | |
|----------------------|---------------------|
| A. hopeful | B. desperate |
| C. disgusting | D. pleased |

7. In 2001, two hijacked airplanes hit and completely destroyed the World Trade Center and nearby buildings in New York, USA. Today, New York City has _____. The subway that runs beneath the WTC buildings has been fixed and starting to work again.

- | | |
|---------------------|--------------------|
| A. separated | B. occupied |
| C. recovered | D. secured |

8. A powerful earthquake has devastated Bam in Iran. Officials _____ the cost of the destruction at around US \$ 100 million. The number of casualties were not officially reported.

- | | |
|---------------------|-----------------------|
| A. estimated | B. requested |
| C. provided | D. accumulated |

9. Many space projects are set up in many countries such as Russia, China, and the US in order to _____ the galaxy and other surrounding planets. Scientists hope that one day they will find other living things out there.

- | | |
|--------------------|---------------------|
| A. relocate | B. cooperate |
| C. explore | D. travel |

10. Besides the regular buses and taxi, the Skytrain and underground railways are _____ means of transportation for commuters and working people in Bangkok. The two projects are set up in order to reduce traffic problems.

- | | |
|----------------------|-----------------------|
| A. obsolete | B. alternative |
| C. compulsory | D. temporary |

Part II: Reading Comprehension (12 items: 12 points)

คำชี้แจง จงอ่านเนื้อเรื่องต่อไปนี้ (Passage I) แล้วตอบคำถาม ข้อ 11-16 โดยวงกลมอักษร A, B, C, หรือ D ที่เป็นคำตอบที่เหมาะสม หากต้องการเปลี่ยนคำตอบ ให้กากบาททับคำตอบเดิม แล้ววงกลมคำตอบใหม่

Passage I

P.1 Spam, the circulation of unwanted electronic messages, is dangerous and expensive for business and individuals and is growing uncontrollably on an epidemic scale. So says an official report prepared for an OECD-EU meeting. Confidence in the entire environment of Internet communications and electronic commerce is at risk. A simple answer is not at hand, says the report that has just been declassified by the OECD. This contrasts to an assertion by Microsoft founder Bill Gates at the Davos Forum that his company would solve the problem within two years.

P.2 Many countries are tightening up legislation, and Internet service companies and direct marketing trade associations are introducing codes of practice. Service suppliers are strengthening defenses and changing **their** data and pricing arrangements, and software makers are developing intelligent filters. Various bodies have begun trying to inform the Internet public of the dangers and defensive **measures**. But the report stresses that legislation alone is not a solution, if indeed there is solution. Even a basic measure whereby people can register as opting out of receiving unwanted mail, and arrangement likely to be ignored by most spammers, is flawed because many users fear that the option itself might be a spam trap.

P.3 Since the Internet is global, and spammers are hard to locate and may work from countries with lax regulations, pursuit of spammers is “extremely difficult”. But organizations are beginning to turn to civil cases for damages and this trend “may deter spammers”, the report suggests. “The major problem of spam is that it creates distrust among Internet users in the digital economy which could have an adverse impact on the development of e-commerce,” the report warns. It is estimated that 51% of corporations “have had a virus disaster”. Viruses in spam have “led to a greater mistrust of e-mail as a secure communication mechanism”, the report says.

P.4 Spam threatens the business environment in multiple ways and is rapidly reaching into the world of mobile telephony, for example by inciting people to call up permanently engaged charge-call numbers. By infiltration techniques such as fishing randomly with thousands of computer-generated variations of known company personal addresses, spammers locate and “harvest” e-mail addresses from the unsuspecting. Such junk mail brings with it many hidden costs for Internet service providers, for their clients, and for business and individuals. It reduces productivity. For example, spam attacks and “paralyzes

or shuts down” a company’s networks. Staff time and company money is consumed in killing unwanted messages and devising counter measures.

P.5 In addition, good mail may be thrown out with the bad, and good direct marketing companies are discredited by bad practice. Spam attacks on company computers and mailing lists pose security and legal risks. Huge number of e-mails may be sent by one distributor at very low cost. “This explains why spam is growing at such an alarming rate”. However, one small study found that more than eight percent of respondents “admitted that they have actually purchased a product promoted by spam.” One anti-spam company, Spamhaus, says that only about 180 known individuals, working in “spam gangs”, are behind all spam received in Europe and the United States. The Japanese telecom operator DoCoMo is one company that has introduced a measure to reduce spam mail, such as charging a premium to originators of waves of e-mails.

(Bangkok Post/February 2, 2004)

คำชี้แจง จงอ่านเนื้อเรื่อง (Passage I) แล้วตอบคำถาม ข้อ 11- 16 โดยวงกลมอักษร A, B, C, หรือ D ที่เป็นคำตอบที่เหมาะสมในกระดาษคำตอบ หากต้องการเปลี่ยนคำตอบ ให้กากบาททับคำตอบเดิม แล้ววงกลมคำตอบใหม่

11. What is the main idea of paragraph 1?
- A. Internet viruses are spreading via the e-mails.
 - B. Unnecessary mail is now causing problems to Internet users.
 - C. There has been an argument between the OECD and Bill Gates.
 - D. Security on the Internet is to be reviewed in the OECD-EU meeting.

12. Which of the following is **NOT** measure to strike back at spam mentioned in paragraph 2?
- A. arrangement cost B. data alteration
C. mail filters developing D. spam trap training
13. What does “**their**” in paragraph 2 refer to?
- A. Defenses of service suppliers
B. Internet service companies
C. Services suppliers
D. Trade associations
14. What does the word “**measures**” in paragraph 2 mean?
- A. reasons B. levels
C. answers D. methods
15. It can be inferred from the passage that _____.
- A. More Internet users will get used to spam.
B. Spammers will not stop attacking Internet services.
C. Spam is a business tool that companies use to compete their rivals.
D. Internet service providers and Internet users do not care about online risks.

16. What can be correct according to paragraph 5?
- A. Spammers also make money by selling products online.
 - B. Spammers are a group of programmers in the United States and Europe.
 - C. The number of unwanted e-mails indicates how clever the spammers are.
 - D. Regular Internet users are believed to be spammers when sending many e-mails at a time.

~~~~~

**คำสั่ง** จงอ่านเนื้อเรื่องต่อไปนี้ (Passage II) แล้วตอบคำถาม ข้อ 17-22 โดยวงกลมอักษร A, B, C, หรือ D ที่เป็นคำตอบที่เหมาะสม หากต้องการเปลี่ยนคำตอบ ให้กากบาททับคำตอบเดิม แล้ววงกลมคำตอบใหม่

### Passage II

P.1 Scientists study it. Doctors recommend it. Millions of people—many of whom do not even own crystals—practice it every day. Why? It is because meditation works. Not only do studies show that meditation can boost the immune system, but brain scans suggest that it may be rewiring the brains to reduce stress. Ten million American adults now say they practice some form of meditation regularly. Meditation classes today are being filled by mainstream Americans who do not own crystals, or subscribe to New Age magazines. And they no longer have to go off to some bearded guru in the woods to do **it**. In fact, it is becoming increasingly hard to avoid meditation. It is offered in schools, hospitals, law firms, government buildings, corporate offices and prisons. There are specially marked meditation rooms in airports alongside the prayer chapels and Internet kiosks. At the Maharishi University schools in Fairfield, Iowa, which include college, high school and elementary classes, the entire elementary school student body meditates together twice daily.

P.2 But the current interest is as much medical as it is cultural. Meditation is being recommended by more and more physicians as a way to prevent, slow or at least control the pain of chronic diseases like heart conditions, AIDS, cancer and infertility. It is also being used to restore balance in the face of such psychiatric disturbances as depression, hyperactivity and attention deficit disorder (ADD). Doctors are embracing meditation not because they think it is hip or cool but because scientific studies are beginning to show that it works, particularly for stress-related conditions. “For 30 years meditation research has told us that it works beautifully as an antidote to stress,” says Daniel Goleman, author of *Destructive Emotions*, a conversation between the Dalai Lama and a group of neuroscientists. “But what’s exciting about the new research is how meditation can train the mind and reshape the brain.” Tests using the most sophisticated imaging techniques suggest that it can actually reset the brain and set the blood boiling.

P.3 The brain, like the body, also undergoes subtle changes during deep meditation. The first scientific studies, in the '60s and '70s, basically proved that meditators are really focused. In India a researcher named B.K. Anand found that yogis could meditate themselves into trances so deep that they did not react when hot test tubes were pressed against their arms. In Japan, a scientist named T. Hirai showed that Zen meditators were so focused on the moment that they never habituated themselves to the sound of a ticking clock (most people eventually block out the noise, but the meditators kept hearing it for hours). Another study showed that master meditators do not flinch at the sound of a gunshot.

P.4 Meanwhile, the evidence from meditation researchers continues to mount. One study, for example, shows that women who meditate and use guided imagery have higher levels of the immune cells known to **combat**

tumors in the breast. This comes after many studies have established that meditation can significantly reduce blood pressure. Given that 60% of doctor visits are the result of stress-related conditions, this is not surprising. Nor is it surprising that meditation can sometimes be used to replace Viagra. But meditation does more than reduce stress, bring harmony and increase focus. Actress Heather Graham started meditating at the suggestion of director David Lynch. “It’s easy to spend a lot of time worrying and obsessing, but meditation puts me in a blissful place,” says Graham. “At the end of the day, all that star stuff doesn’t mean anything. Transcendental Meditation reminds you that it’s how you feel inside that’s important. If you have that, you have everything.”

(TIME, August 04, 2003)

**คำสั่ง** จงอ่านเนื้อเรื่อง (Passage II) แล้วตอบคำถาม ข้อ 17 – 22 โดยวงกลมอักษร A, B, C, หรือ D ที่เป็นคำตอบที่เหมาะสมในกระดาษคำตอบ หากต้องการเปลี่ยนคำตอบ ให้กากบาททับคำตอบเดิม แล้ววงกลมคำตอบใหม่

17. What does “it” in paragraph 1 refer to?
- |                          |                          |
|--------------------------|--------------------------|
| A. going to the forest   | B. buying magazines      |
| C. practicing meditation | D. studying about stress |
18. What is **NOT** a result from recent studies about meditation as mentioned in paragraph 2? Meditation can \_\_\_\_\_.
- |                                           |
|-------------------------------------------|
| A. prevent heart and eye disease.         |
| B. reduce stress-related health problems. |
| C. train our mind and modify the brain.   |
| D. lessen pain caused by cancers.         |

19. What is the main idea of paragraph 3?
- A. A secret power of meditation and its relationship with Asian people
  - B. Evidence of changes in the brain and body during meditation
  - C. Activities that usually disturb people while practicing meditation
  - D. Levels of meditation that monks in Asian countries can achieve
20. What does “**combat**” in paragraph 4 mean?
- A. maintain
  - B. protect
  - C. remove
  - D. resist
21. What is Heather Graham’s opinion about meditation?
- A. Practicing meditation helped her become a famous actress.
  - B. Meditation makes her feel at peace and her mind clear.
  - C. Practicing meditation brings her a new friend, David Lynch.
  - D. Meditation helps improve her eyesight.
22. What can we infer from the passage?
- A. People will spend more time on meditation but not on their work.
  - B. More and more young persons will practice meditation because they want to be famous.
  - C. The majority of the people around the world meditate in order to avoid communicating with others.
  - D. Studies and evidence about health benefits from doing meditation will increase.

~~~ End of the Paper-Pencil Test ~~~

Thank You Very Much for Your Cooperation

Appendix B

Reading Comprehension Test of English (RCTE) (The Computer Mode)



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

| Test Series | Stem | A | B | C | D | Correct Choice |
|-------------|--|-------------|--------------|--------------|------------|----------------|
| 1 | Moviemakers usually invest a lot of money in advertising their movies in order to _____ them up and attract the public attention as much as possible. | clean | call | hype | mount | C |
| 2 | To survive in the competitive markets, entrepreneurs need to be _____ instead of copying each other. For example, silk and hand woven clothes should be made wrinkle-free. | adventurous | powerful | successful | innovative | D |
| 3 | The trend for form-fitting jeans and midriff-exposing tops shows no signs of stopping and neither does the obsession with the hourglass _____ needed to wear such clothes among female Thai teenagers. | outfit | model | figure | statue | C |
| 4 | Internet chat sites in the United Kingdom were ordered to shut down since young girls were sexually _____ after they arranged to meet men they met in the chat rooms. | described | abused | bewitched | criticized | B |
| 5 | A brand new and fun-to-use smartly designed camera phone with a lot of useful _____ including Bluetooth wireless connectivity, a built-in modem for e-mailing and Internet browsing will hit the shelves in the spring. | prototypes | arrangements | compartments | features | D |
| 6 | Many well-known persons and business executives attended a charity jazz concert that was set up to raise funds to help villagers in a rural village build a public library. Some of them _____ money and books to the head of the village. | withdrew | allowed | contributed | divided | C |

| Test Series | Stem | A | B | C | D | Correct Choice |
|-------------|---|------------|-------------|-----------|-------------|----------------|
| 7 | In some criminal cases, police refuse to _____ details of the victims and suspects' identity. This is to protect their rights according to the constitution. | release | execute | combine | investigate | A |
| 8 | The landing of the US robotic probe--Spirit-- _____ when it beamed panoramic color images of Mars surface back to Earth. The craft landed on Mars landscape on target where scientists believe may be the site of a dry lake. | improved | succeeded | postponed | aborted | B |
| 9 | Scientists have found a new type of lawn that we do not have to mow very often. They have genetically developed a grass that is _____. It is resistant to drought and can return itself from dry tangled form to green after receiving only a few drops of water. | beneficial | basic | harmless | resilient | D |
| 10 | Spices and herbal ingredients used in Thai cuisine do not only make the food tasty but also work as _____ for illnesses. | remedy | facilitator | stimuli | assistant | A |

The Vocabulary Section (continued)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

| Test Series | Stem | A | B | C | D | Correct Choice |
|-------------|--|---|---|---|---|----------------|
| 1 | What does “supplies” in paragraph 1 mean? | belongings | services | materials | ornaments | C |
| 2 | What can be inferred from paragraph 1? | Handheld devices will probably be accepted and used in schools. | American teenagers will feel ashamed for not having handheld devices. | Students at Palo Alto High School are given handheld computers by the US Department of Education. | Many schools are competing against each other about the number of handheld devices in their school. | A |
| 3 | What is NOT one of the handheld devices mentioned in paragraph 2? | Laptops | Pagers | Organizers | PDAs | B |
| 4 | If high-tech devices with wireless Internet access are allowed in class, it is very likely that _____. | students will lose interest in printed texts. | students will use the devices to cheat on tests. | teachers will have to teach extra classes on Internet skills. | teachers will have to check notes that students make in class. | B |
| 5 | What does paragraph 3 mainly discuss? | Properties and qualities of the handheld devices | Availability and increasing use of the handheld devices | Activities for which students use the handheld devices | Regulations to control the use of the handheld devices | D |
| 6 | In paragraph 3, what does “there” refer to? | Palo Alto High School | Schools in the high-tech region | East Dubuque High School | Schools in other states | C |

| Test Series | Stem | A | B | C | D | Correct Choice |
|-------------|---|---|---|---|--|----------------|
| 1 | According to paragraph 1, spas in Europe nowadays are_____. | initiated by European people who live in the US. | using methods by which the Roman people utilized water. | created by and belong only to Greek and Roman people. | situated in the historical areas of Greek and Roman empires. | B |
| 2 | What does the word “soothing” in paragraph 2 mean? | healing | disturbing | inducing | protecting | A |
| 3 | What can we infer from paragraph 2? | Modern medical treatments and doctors have also been influenced by Kneipp’s idea. | Spas in Germany are established to provide services only to people with certain diseases. | Water applications for ailments have just originated about a hundred years ago. | Therapeutic uses of water are well accepted among people with health problems especially in Germany. | D |
| 4 | What are NOT used when doing cold body wraps? | dry blankets | cold and hot water | soaked sheets | salt and mud | D |
| 5 | What does “this” in paragraph 3 refer to? | the body wrapped in a soaked sheet | the body | a sheet soaked in cold water | cold water | A |
| 6 | What is the main idea of paragraph 4? | Health benefits from using cold baths | Recommendations of applying hot water to the body | Advantages and warnings of water healing treatments | Health problems that can be caused by the use of hot and cold water | C |

Appendix C

The Description of the Reading Comprehension Test of English (RCTE)

| Section | Vocabulary | Technology/Internet | | Alternative Therapy | |
|---|---|---|--|--|--|
| | | Reading I | Reading II | Reading III | Reading IV |
| | | Online Epidemic
(Bangkok Post. February 02, 2004) | “No Calls in Class, Kids”
(Nation. October 23, 2003) | Meditation
(TIME. August 04, 2003) | Naturopathy
(Phil, Mark E.B. 1996. The Guide to Natural Therapies. London: Lorenz Books) |
| Readability level | | 12.0 | 10.7 | 11.1 | 12 |
| Reading ease
(Flesch-Kincaid Reading Ease formula) | | 38.5 | 53.2 | 47.5 | 39.1 |
| Mode of presentation | P&P and CBT | P&P | CBT | P&P | CBT |
| Test Items | 40
(P&P = 20, CBT = 20) | 15 | 15 | 15 | 15 |
| Test formats | MC | MC | MC | MC | MC |
| Aim of the section | To measure English grammar, English language use, textual and pragmatic knowledge | To measure reading comprehension skills: | | | |
| | | <ul style="list-style-type: none"> Scanning Skimming Identifying pronoun references Guessing word meaning from the context Identifying the topic and main idea | | <ul style="list-style-type: none"> Recognizing supporting details Summarizing information Making inferences Drawing conclusion | |
| Type of passages/texts | Descriptive, argumentative | Argumentative | Argumentative | Descriptive | Descriptive |
| Length | 3-4 statements in each item | 450-500 words each | | | |

Proportion of the Test Items on Each Reading Skill

| Bloom's Taxonomy | | Reading comprehension skills | |
|---------------------|----------|--|----------|
| Knowledge level | 16 items | Pronoun reference | 6 items |
| | | Finding specific information | 10 items |
| Comprehension level | 9 items | Understanding explicit and implied information | 9 items |
| | | Interpret ideas | |
| Analysis level | 8 items | Making inferences, identifying components | 8 items |
| Synthesis level | 27 items | Finding main idea | 9 items |
| | | Finding topic/title | 1 items |
| | | Guessing word meaning | 17 items |
| Total | 60 items | Total | 60 items |

Reading Comprehension Skill Analysis Based on Bloom's Taxonomy

| Title | Bloom's Taxonomy (Competence) | | Reading Skills | |
|--|-------------------------------|------------------------------------|------------------------------------|---|
| Online Epidemic (P&P) | Knowledge | 5 | Finding specific information | 3 |
| | | | Finding pronoun references | 2 |
| | Comprehension | 3 | Understand information | 3 |
| | Analysis | 1 | Making inference | 1 |
| | Synthesis | 6 | Guessing word meaning from context | 4 |
| | | | Finding main idea & title | 2 |
| Total | 15 | | | |
| “No Calls in Class, Kids” (CBT) | Knowledge | 3 | Finding specific information | 2 |
| | | | Finding pronoun references | 1 |
| | Comprehension | 2 | Understanding implied idea | 1 |
| | | | Interpreting implied idea | 1 |
| | Analysis | 3 | Making inferences | 3 |
| | Synthesis | 7 | Guessing word meaning from context | 5 |
| Finding main idea | | | 2 | |
| Total | 15 | | | |
| Naturopathy (P&P) | Knowledge | 3 | Finding specific information | 2 |
| | | | Finding pronoun reference | 1 |
| | Comprehension | 3 | Interpreting implied idea | 1 |
| | | | Understanding ideas | 2 |
| | Analysis | 2 | Making inferences | 1 |
| | | | Drawing conclusion | 1 |
| Synthesis | 7 | Guessing word meaning from context | 4 | |
| | | Finding main idea | 3 | |
| Total | 15 | | | |
| Meditation (CBT) | Knowledge | 4 | Finding specific information | 2 |
| | | | Finding pronoun references | 2 |
| | Comprehension | 1 | Understanding ideas | 1 |
| | Analysis | 3 | Making inferences | 2 |
| | | | Identifying components | 1 |
| | Synthesis | 7 | Guessing word meaning from context | 4 |
| Finding main idea | | | 2 | |
| Finding title | | | 1 | |
| Total | 15 | | | |

Appendix D



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Pilot Study Item Analysis (The Paper-Pencil Mode: 50 Test Items)

| Item No. | Difficulty index | Delta | Discrimination index | Biserial correlation coefficients (r_{bis}) |
|---------------|------------------|-------|----------------------|---|
| Vocabulary 1* | 0.286 | 15.30 | 0.333 | 0.544 |
| 2 | 0.029 | 20.65 | -0.111 | -0.367 |
| *3 | 0.629 | 11.65 | 0.667 | 0.553 |
| 4 | 0.343 | 14.66 | 0.111 | -0.106 |
| *5 | 0.257 | 15.65 | 0.556 | 0.690 |
| 6 | 0.257 | 15.65 | 0.000 | 0.195 |
| 7 | 0.114 | 17.86 | 0.000 | 0.047 |
| *8 | 0.400 | 14.05 | 0.333 | 0.309 |
| 9 | 0.686 | 11.03 | 0.556 | 0.380 |
| 10 | 0.629 | 11.65 | 0.889 | 0.742 |
| *11 | 0.171 | 16.83 | 0.222 | 0.564 |
| 12 | 0.571 | 12.24 | 0.667 | 0.531 |
| 13 | 0.686 | 11.03 | 0.333 | 0.474 |
| *14 | 0.229 | 16.04 | 0.444 | 0.528 |
| *15 | 0.629 | 11.65 | 0.222 | 0.503 |
| *16 | 0.543 | 12.53 | 0.778 | 0.743 |
| *17 | 0.686 | 11.03 | 0.444 | 0.527 |
| *18 | 0.514 | 12.82 | 0.444 | 0.376 |
| *19 | 0.171 | 16.83 | 0.444 | 0.751 |
| *20 | 0.314 | 14.97 | 0.444 | 0.550 |
| *RC 21 | 0.486 | 13.18 | 0.778 | 0.628 |
| 22 | 0.171 | 16.83 | -0.333 | -0.426 |
| *23 | 0.343 | 14.66 | 0.444 | 0.509 |
| *24 | 0.657 | 11.34 | 0.444 | 0.446 |
| *25 | 0.457 | 13.47 | 0.444 | 0.388 |
| *26 | 0.371 | 14.35 | 0.111 | -0.035 |
| 27 | 0.257 | 15.65 | 0.111 | 0.253 |
| 28 | 0.286 | 15.30 | 0.333 | 0.380 |
| 29 | 0.457 | 13.47 | -0.111 | -0.230 |
| 30 | 0.286 | 15.30 | 0.333 | 0.282 |
| *31 | 0.229 | 16.01 | 0.556 | 0.712 |
| 32 | 0.343 | 14.66 | 0.333 | 0.324 |
| 33 | 0.371 | 14.35 | 0.333 | 0.191 |
| 34 | 0.257 | 15.65 | 0.444 | 0.529 |
| 35 | 0.486 | 13.18 | 0.111 | 0.281 |
| 36 | 0.714 | 10.70 | 0.556 | 0.638 |
| *37 | 0.714 | 10.70 | 0.778 | 0.780 |
| 38 | 0.629 | 11.65 | 0.444 | 0.583 |
| 39 | 0.371 | 14.35 | 0.222 | 0.204 |
| 40 | 0.543 | 12.53 | 0.444 | 0.259 |
| 41 | 0.686 | 11.03 | 0.556 | 0.518 |
| *42 | 0.371 | 14.35 | 0.667 | 0.552 |
| 43 | 0.543 | 12.53 | 0.778 | 0.810 |
| 44 | 0.286 | 15.30 | -0.222 | -0.285 |
| *45 | 0.286 | 15.30 | 0.222 | 0.315 |
| 46 | 0.171 | 16.83 | 0.222 | 0.507 |
| *47 | 0.371 | 14.35 | 0.222 | 0.383 |
| *48 | 0.743 | 10.35 | 0.222 | 0.395 |
| *49 | 0.475 | 13.47 | 0.667 | 0.611 |
| 50 | 0.629 | 11.65 | 0.222 | 0.274 |

K-R 20 and Alpha = 0.826 (Note: * are the selected items for the main study)

Pilot Study Item Analysis (Computer Mode: 50 Test Items)

| Item No. | Difficulty index | Delta | Discrimination index | Biserial correlation coefficients (r_{bis}) |
|--------------|------------------|-------|----------------------|---|
| Vocabulary 1 | 0.829 | 9.17 | 0.444 | 0.623 |
| *2 | 0.286 | 15.30 | 0.556 | 0.452 |
| 3 | 0.171 | 16.83 | 0.000 | 0.006 |
| *4 | 0.629 | 11.65 | 0.667 | 0.765 |
| 5 | 0.400 | 14.05 | 0.111 | 0.329 |
| *6 | 0.086 | 18.51 | 0.111 | 0.280 |
| *7 | 0.486 | 13.18 | 0.333 | 0.312 |
| *8 | 0.429 | 13.76 | 0.778 | 0.691 |
| *9 | 0.486 | 13.18 | 0.222 | 0.229 |
| 10 | 0.257 | 15.65 | 0.111 | 0.191 |
| 11 | 0.800 | 9.59 | 0.333 | 0.549 |
| 12 | 0.314 | 14.97 | 0.222 | 0.117 |
| *13 | 0.429 | 13.76 | 0.333 | 0.415 |
| 14 | 0.629 | 11.65 | 0.333 | 0.379 |
| 15 | 0.657 | 11.34 | 0.333 | 0.275 |
| 16 | 0.286 | 15.30 | 0.111 | 0.189 |
| *17 | 0.571 | 12.24 | 0.667 | 0.596 |
| 18 | 0.543 | 12.53 | 0.444 | 0.611 |
| 19 | 0.800 | 9.59 | 0.444 | 0.572 |
| 20 | 0.571 | 12.24 | 0.778 | 0.646 |
| RC 21 | 0.371 | 14.35 | 0.556 | 0.496 |
| *22 | 0.400 | 14.05 | 0.444 | 0.363 |
| *23 | 0.543 | 12.53 | 0.667 | 0.578 |
| *24 | 0.686 | 11.03 | 0.444 | 0.505 |
| 25 | 0.743 | 10.35 | 0.222 | 0.248 |
| *26 | 0.514 | 12.82 | 0.778 | 0.819 |
| 27 | 0.429 | 13.76 | 0.333 | 0.373 |
| *28 | 0.543 | 12.53 | 0.667 | 0.528 |
| 29 | 0.257 | 15.65 | 0.333 | 0.359 |
| *30 | 0.486 | 13.18 | 0.778 | 0.670 |
| 31 | 0.657 | 11.34 | 0.444 | 0.425 |
| 32 | 0.600 | 11.95 | 0.667 | 0.613 |
| 33 | 0.086 | 18.51 | 0.222 | 0.547 |
| 34 | 0.686 | 11.03 | 0.000 | 0.117 |
| 35 | 0.314 | 14.97 | 0.222 | 0.162 |
| 36 | 0.857 | 8.28 | 0.556 | 0.917 |
| 37 | 0.486 | 13.18 | 0.778 | 0.786 |
| 38 | 0.314 | 14.97 | 0.444 | 0.477 |
| *39 | 0.457 | 13.47 | 0.778 | 0.596 |
| 40 | 0.400 | 14.05 | 0.333 | 0.431 |
| *41 | 0.486 | 13.18 | 0.444 | 0.512 |
| *42 | 0.543 | 12.53 | 0.556 | 0.486 |
| 43 | 0.514 | 20.65 | -0.111 | -0.537 |
| 44 | 0.171 | 16.83 | 0.000 | 0.179 |
| *45 | 0.543 | 12.53 | 0.889 | 0.661 |
| *46 | 0.371 | 14.35 | 0.333 | 0.273 |
| 47 | 0.457 | 13.47 | 0.444 | 0.505 |
| *48 | 0.486 | 13.18 | 0.667 | 0.645 |
| 49 | 0.343 | 14.66 | 0.222 | 0.415 |
| 50 | 0.457 | 13.47 | 0.667 | 0.480 |

K-R 20 and Alpha = 0.869 (Note: * are the selected items for the main study)

Statistic Values of the Selected Vocabulary Items

| Item No. | Correct choice | Difficulty index | Delta | Discrimination index | Biserial Correlation Coefficient/ r_{bis} |
|----------------|----------------|------------------|-------|----------------------|---|
| P&P | | | | | |
| 1 | C/3 | 0.286 | 15.30 | 0.333 | 0.544 |
| 3 | D/4 | 0.629 | 11.65 | 0.667 | 0.553 |
| 5 | D/4 | 0.257 | 15.65 | 0.556 | 0.690 |
| 8 | A/1 | 0.400 | 14.05 | 0.333 | 0.309 |
| 14 | B/2 | 0.229 | 16.01 | 0.444 | 0.528 |
| 19 | D/4 | 0.171 | 16.83 | 0.444 | 0.751 |
| 20 | A/1 | 0.314 | 14.97 | 0.444 | 0.550 |
| 11 | D/4 | 0.171 | 16.83 | 0.222 | 0.564 |
| 15 | C/3 | 0.629 | 11.65 | 0.222 | 0.503 |
| 16 | D/4 | 0.543 | 12.53 | 0.778 | 0.743 |
| 17 | C/3 | 0.686 | 11.03 | 0.444 | 0.527 |
| 18 | B/2 | 0.514 | 12.82 | 0.444 | 0.376 |
| CBT | | | | | |
| 2 | C/3 | 0.286 | 15.30 | 0.556 | 0.452 |
| 4 | D/4 | 0.629 | 11.65 | 0.667 | 0.765 |
| 6 | C/3 | 0.314 | 14.97 | 0.333 | 0.468 |
| 7 | B/2 | 0.486 | 13.18 | 0.333 | 0.312 |
| 8 | D/4 | 0.429 | 13.76 | 0.778 | 0.691 |
| 9 | C/3 | 0.486 | 13.18 | 0.222 | 0.229 |
| 13 | C/3 | 0.429 | 13.76 | 0.333 | 0.415 |
| 17 | C/3 | 0.571 | 12.24 | 0.667 | 0.596 |

**Statistic Values of the Selected Items in the Reading Comprehension Section
(4 passages)**

| | Item No. | Correct choice | Difficulty index | Delta | Discrimination Index | Biserial Correlation Coefficient/ r_{bis} |
|-----------------------------------|-------------------------------|----------------|------------------|-------|----------------------|---|
| Scanning | P&P/Online
23 | D/4 | 0.343 | 14.66 | 0.444 | 0.509 |
| | CBT/NoCalls
24 | B/2 | 0.686 | 11.03 | 0.444 | 0.505 |
| | P&P/Medittn.
42 | C/3 | 0.371 | 14.35 | 0.667 | 0.552 |
| | CBT/Natpy.
45 | C/3 | 0.543 | 12.53 | 0.889 | 0.661 |
| Finding pronoun references | P&P/Spam
24 | C/3 | 0.657 | 11.34 | 0.444 | 0.446 |
| | CBT/NoCalls
30 | C/3 | 0.486 | 13.18 | 0.778 | 0.670 |
| | P&P/Medttt.
37 | C/3 | 0.714 | 10.70 | 0.778 | 0.780 |
| | CBT/Natpy.
46 | A/1 | 0.371 | 14.35 | 0.333 | 0.273 |

**Statistic Values of the Selected Items in the Reading Comprehension Section
(Cont'd)**

| | Item No. | Correct choice | Difficulty index | Delta | Discrimination index | Biserial Correlation Coefficient/ r_{bis} |
|--|----------------------------|----------------|------------------|-------|----------------------|---|
| Guessing word meaning from context | P&P/Spam 25 | D/4 | 0.457 | 13.47 | 0.444 | 0.388 |
| | CBT/NoCalls 22 | C/3 | 0.400 | 14.05 | 0.444 | 0.363 |
| | P&P/Medtttn. 47 | D/4 | 0.371 | 14.35 | 0.222 | 0.383 |
| | CBT/Natpy. 41 | A/1 | 0.486 | 13.18 | 0.444 | 0.512 |
| Understanding information/comprehension | P&P/Spam 31 | A/1 | 0.229 | 16.01 | 0.556 | 0.712 |
| | CBT/NoCalls 26 | B/2 | 0.514 | 12.82 | 0.778 | 0.819 |
| | P&P/Medtttn. 48 | B/2 | 0.743 | 10.35 | 0.222 | 0.395 |
| | CBT/Natpy. 39 | B/2 | 0.457 | 13.47 | 0.778 | 0.596 |
| Finding main ideas | P&P/Spam 21 | B/2 | 0.486 | 13.18 | 0.778 | 0.628 |
| | CBT/NoCalls 28 | B/2 | 0.543 | 12.53 | 0.667 | 0.528 |
| | P&P/Medtttn. 45 | B/2 | 0.286 | 15.30 | 0.222 | 0.315 |
| | CBT/Natpy. 48 | C/3 | 0.486 | 13.18 | 0.667 | 0.645 |
| Making inferences | P&P/Spam 26 | B/2 | 0.371 | 14.35 | 0.111 | -0.035 |
| | CBT/NoCalls 23 | A/1 | 0.543 | 12.53 | 0.667 | 0.578 |
| | P&P/Medtttn. 49 | D/4 | 0.457 | 13.47 | 0.667 | 0.661 |
| | CBT/Natpy. 42 | B/2 | 0.543 | 12.53 | 0.556 | 0.486 |

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Statistic Values of Items in the Main Study

Vocabulary

| Item No. | Correct choice | Difficulty index | Delta | Discrimination index | Biserial Correlation Coefficient/ r_{bis} |
|----------------|----------------|------------------|-------|----------------------|---|
| P&P | | | | | |
| 1 | C/3 | 0.303 | 15.10 | 0.412 | 0.539 |
| 2 | D/4 | 0.576 | 12.20 | 0.412 | 0.410 |
| 3 | B/2 | 0.348 | 14.60 | 0.417 | 0.519 |
| 4 | A/1 | 0.364 | 14.43 | 0.529 | 0.462 |
| 5 | D/4 | 0.136 | 17.43 | 0.235 | 0.479 |
| 6 | B/2 | 0.242 | 15.83 | 0.588 | 0.680 |
| 7 | C/3 | 0.500 | 13.00 | 0.412 | 0.392 |
| 8 | A/2 | 0.561 | 12.35 | 0.471 | 0.575 |
| 9 | C/3 | 0.773 | 9.97 | 0.412 | 0.432 |
| 10 | B/2 | 0.500 | 13.00 | 0.235 | 0.269 |
| CBT | | | | | |
| 1 | C/3 | 0.167 | 16.91 | 0.235 | 0.333 |
| 2 | D/4 | 0.485 | 13.19 | 0.765 | 0.682 |
| 3 | C/3 | 0.318 | 14.93 | 0.294 | 0.192 |
| 4 | B/2 | 0.455 | 13.50 | 0.471 | 0.505 |
| 5 | D/4 | 0.561 | 12.35 | 0.294 | 0.318 |
| 6 | C/3 | 0.500 | 13.00 | 0.529 | 0.479 |
| 7 | A/1 | 0.485 | 13.19 | 0.471 | 0.550 |
| 8 | B/2 | 0.530 | 12.66 | 0.529 | 0.591 |
| 9 | D/4 | 0.136 | 17.43 | 0.353 | 0.764 |
| 10 | A/1 | 0.424 | 13.80 | 0.882 | 0.797 |

Statistic Values of the Reading Comprehension Items in the Main Study

| | Item No. | Correct choice | Difficulty index | Delta | Discrimination Index | Biserial Correlation Coefficient/ r_{bis} |
|-----------------------------------|-------------------------------|----------------|------------------|-------|----------------------|---|
| Scanning | P&P/Online
12 | D/4 | 0.348 | 14.60 | 0.471 | 0.548 |
| | CBT/NoCalls
3 | B/2 | 0.652 | 11.40 | 0.235 | 0.317 |
| | P&P/Medittn.
18 | A/1 | 0.242 | 15.83 | 0.765 | 0.923 |
| | CBT/Natpy.
4 | D/4 | 0.439 | 13.65 | 0.353 | 0.418 |
| Finding pronoun references | P&P/Spam
13 | C/3 | 0.621 | 11.73 | 0.471 | 0.446 |
| | CBT/NoCalls
6 | C/3 | 0.394 | 14.12 | 0.294 | 0.288 |
| | P&P/Medtttn.
17 | C/3 | 0.682 | 11.07 | 0.588 | 0.686 |
| | CBT/Natpy.
5 | A/1 | 0.485 | 13.19 | 0.353 | 0.400 |

Cont'd

| | Item No. | Correct choice | Difficulty index | Delta | Discrimination index | Biserial Correlation Coefficient/ r_{bis} |
|---|----------------|----------------|------------------|-------|----------------------|---|
| Guessing word meaning from context | P&P/Spam 14 | D/4 | 0.515 | 12.81 | 0.294 | 0.357 |
| | CBT/NoCalls 1 | C/3 | 0.273 | 15.46 | 0.471 | 0.573 |
| | P&P/Medttn. 20 | D/4 | 0.258 | 15.64 | 0.176 | 0.224 |
| | CBT/Natpy. 2 | A/1 | 0.424 | 13.80 | 0.529 | 0.474 |
| Understanding information/comprehension | P&P/Spam 16 | A/1 | 0.258 | 15.64 | 0.412 | 0.550 |
| | CBT/NoCalls 4 | B/2 | 0.515 | 12.81 | 0.529 | 0.474 |
| | P&P/Medttn. 21 | B/2 | 0.712 | 10.72 | 0.471 | 0.563 |
| | CBT/Natpy. 1 | B/2 | 0.561 | 12.35 | 0.235 | 0.309 |
| Finding main ideas | P&P/Spam 11 | B/2 | 0.379 | 14.27 | 0.647 | 0.624 |
| | CBT/NoCalls 5 | D/4 | 0.424 | 13.80 | 0.588 | 0.616 |
| | P&P/Medttn. 19 | B/2 | 0.485 | 13.19 | 0.235 | 0.134 |
| | CBT/Natpy. 6 | C/3 | 0.379 | 14.27 | 0.706 | 0.761 |
| Making inferences | P&P/Spam 15 | B/2 | 0.273 | 15.46 | 0.529 | 0.538 |
| | CBT/NoCalls 2 | A/1 | 0.530 | 12.66 | 0.706 | 0.600 |
| | P&P/Medttn. 22 | C/3 | 0.636 | 11.57 | 0.706 | 0.692 |
| | CBT/Natpy. 3 | D/4 | 0.439 | 13.65 | 0.353 | 0.182 |

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix E

The Attitude Questionnaire (Pilot Study)

แบบสอบถามทัศนคติต่อการทดสอบสัมฤทธิ์ผลการอ่านภาษาอังกฤษด้วยคอมพิวเตอร์
และแบบประเพณีนิยม (กระดาษ)

ข้อมูลของผู้ตอบแบบสอบถาม

รหัสนักศึกษา คะแนนสอบเอ็นทรานซ์/โควต้าวิชาภาษาอังกฤษคะแนน
Course 103 104 203 Section อายุ ปี เพศ หญิง ชาย

คำชี้แจง

แบบสอบถามนี้แบ่งออกเป็น 4 ตอน มีทั้งหมด 4 หน้า ให้นักศึกษาอ่านข้อความเกี่ยวกับการทดสอบสัมฤทธิ์ผลการอ่านภาษาอังกฤษด้วยคอมพิวเตอร์และแบบประเพณีนิยมในแต่ละส่วน แล้วเลือกกาบาทหมายเลขที่ตรงกับความคิดเห็นของนักศึกษาที่มีต่อข้อความนั้นๆ ในการตอบแบบสอบถามนี้ไม่มีการตีความความคิดเห็นใดๆ ถูกหรือผิด ขอความกรุณานักศึกษาตอบแบบสอบถามทุกข้อด้วย ข้อมูลที่ได้จากแบบสอบถามนี้จะนำไปใช้ในการพัฒนาการทดสอบด้านการใช้ภาษาอังกฤษด้วยคอมพิวเตอร์

ทั้งนี้ ผู้ออกแบบสอบถามขอขอบคุณนักศึกษาทุกคนที่ช่วยตอบแบบสอบถาม และขอให้สัญญาว่าข้อมูลส่วนบุคคลต่างๆที่ปรากฏในแบบสอบถามนี้จะไม่ถูกเปิดเผยแต่อย่างใด

ตัวอย่างการตอบแบบสอบถาม

ให้นักศึกษาอ่านข้อความแล้วเลือกกาบาทหมายเลขที่ตรงกับระดับความคิดเห็นของนักศึกษาที่มีต่อข้อความนั้นๆ โดยที่

- | | | |
|-----------|---------------|--|
| หมายเลข 1 | มีความหมายว่า | นักศึกษามีความเห็นด้วยน้อยหรือไม่เห็นด้วยมาก |
| หมายเลข 2 | มีความหมายว่า | นักศึกษาไม่เห็นด้วย |
| หมายเลข 3 | มีความหมายว่า | นักศึกษาไม่มีความคิดเห็นในประเด็นนี้ |
| หมายเลข 4 | มีความหมายว่า | นักศึกษามีความเห็นด้วย |
| หมายเลข 5 | มีความหมายว่า | นักศึกษามีความเห็นด้วยมาก |

| | | | | | |
|---|-------------------------------------|---|---|---|---|
| ก. ในอนาคต คอมพิวเตอร์จะมาแทนที่ครูผู้สอน | <input checked="" type="checkbox"/> | 2 | 3 | 4 | 5 |
|---|-------------------------------------|---|---|---|---|

หากนักศึกษาต้องการเปลี่ยน ให้นักศึกษาขีดทับตัวเลือกเดิม แล้วกาบาทหมายเลขใหม่ ดังนี้

| | | | | | |
|---|-------------------------------------|---|---|-------------------------------------|---|
| ก. ในอนาคต คอมพิวเตอร์จะมาแทนที่ครูผู้สอน | <input checked="" type="checkbox"/> | 2 | 3 | <input checked="" type="checkbox"/> | 5 |
|---|-------------------------------------|---|---|-------------------------------------|---|

ตอนที่ 1 ทักษะคิดเกี่ยวกับ ข้อสอบภาษาอังกฤษ โดยทั่วไป

หมายเลข 1 ----ไม่เห็นด้วยมาก หมายเลข 2 ----ไม่เห็นด้วย หมายเลข 3 ----ไม่มีความคิดเห็น
 หมายเลข 4 ----เห็นด้วย หมายเลข 5 ----เห็นด้วยมาก

| | | | | | |
|--|---|---|---|---|---|
| 1. นักศึกษาไม่รู้สึกริวกัดกังวลเวลาทำข้อสอบวิชาภาษาอังกฤษ | 1 | 2 | 3 | 4 | 5 |
| 2. เวลาทำข้อสอบวิชาภาษาอังกฤษ นักศึกษามักคิดบ่อยๆว่าข้อสอบนั้นยาก | 1 | 2 | 3 | 4 | 5 |
| 3. ข้อสอบภาษาอังกฤษมีคำศัพท์ใหม่ทำให้ไม่อยากทำข้อสอบนั้น | 1 | 2 | 3 | 4 | 5 |
| 4. เนื้อเรื่องในข้อสอบมีความน่าสนใจ | 1 | 2 | 3 | 4 | 5 |
| 5. นักศึกษาสามารถใช้ทักษะการอ่านภาษาอังกฤษที่เรียนมาทำข้อสอบวิชาภาษาอังกฤษ | 1 | 2 | 3 | 4 | 5 |
| 6. เวลาที่นักศึกษาทำข้อสอบ นักศึกษาจะตั้งใจจดอยู่กับข้อสอบ | 1 | 2 | 3 | 4 | 5 |
| 7. นักศึกษามีการเตรียมตัวเป็นอย่างดีก่อนทำข้อสอบ | 1 | 2 | 3 | 4 | 5 |
| 8. นักศึกษาไม่เข้าใจคำถามในข้อสอบภาษาอังกฤษเลย | 1 | 2 | 3 | 4 | 5 |
| 9. นักศึกษาเข้าใจเนื้อเรื่องในข้อสอบภาษาอังกฤษ | 1 | 2 | 3 | 4 | 5 |

ตอนที่ 2 ประสิทธิภาพการใช้คอมพิวเตอร์

หมายเลข 1----ไม่เคยใช้เลย หรือใช้ไม่เป็นเลย หมายเลข 2 ----ใช้น้อยกว่า 3 ครั้งต่อสัปดาห์
 หมายเลข 3 ---- ใช้ ครั้ง 3-4 ต่อสัปดาห์ หรือ พอใช้ได้ หรือ ใช้ได้บ้าง
 หมายเลข 5 -----ใช้ทุกวัน หรือ ใช้ได้ดีมาก หมายเลข 4 ----ใช้เกือบทุกวัน หรือ ใช้ได้ดี

| | | | | | |
|---|---|---|---|---|---|
| 1. นักศึกษาใช้คอมพิวเตอร์ที่บ้าน | 1 | 2 | 3 | 4 | 5 |
| 2. นักศึกษาใช้คอมพิวเตอร์ที่มหาวิทยาลัย | 1 | 2 | 3 | 4 | 5 |
| 3. นักศึกษาใช้คอมพิวเตอร์ที่ร้านอินเทอร์เน็ต | 1 | 2 | 3 | 4 | 5 |
| 4. นักศึกษาใช้คอมพิวเตอร์พิมพ์รายงานเป็นภาษาไทย | 1 | 2 | 3 | 4 | 5 |
| 5. นักศึกษาใช้คอมพิวเตอร์พิมพ์รายงานเป็นภาษาอังกฤษ | 1 | 2 | 3 | 4 | 5 |
| 6. นักศึกษาใช้คอมพิวเตอร์สร้างกราฟ วาดรูป สร้างงานนำเสนอ | 1 | 2 | 3 | 4 | 5 |
| 7. นักศึกษาใช้ “mouse” หรือ “touch pad” | 1 | 2 | 3 | 4 | 5 |
| 8. นักศึกษาใช้คอมพิวเตอร์เพื่อท่องอินเทอร์เน็ต | 1 | 2 | 3 | 4 | 5 |
| 9. นักศึกษาใช้คอมพิวเตอร์เพื่อค้นหาข้อมูลจากเว็บไซต์ทางการศึกษา | 1 | 2 | 3 | 4 | 5 |

ตอนที่ 3 ทศนคติเกี่ยวกับข้อสอบภาษาอังกฤษแบบประเพณีนิยม (กระดาษ) และด้วยคอมพิวเตอร์

หมายเลข 1 -----ไม่เห็นด้วยมาก หมายเลข 2 -----ไม่เห็นด้วย หมายเลข 3 -----ไม่มีความคิดเห็น
 หมายเลข 4 ----- เห็นด้วย หมายเลข 5 ----- เห็นด้วยมาก

| | | | | | |
|---|---|---|---|---|---|
| 1. นักศึกษารู้สึกสบายใจที่จะทำข้อสอบวิชาภาษาอังกฤษแบบประเพณีนิยม (กระดาษ) | 1 | 2 | 3 | 4 | 5 |
| 2. นักศึกษารู้สึกชอบและอยากทำข้อสอบด้วยคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |
| 3. นักศึกษาคิดว่าเนื้อเรื่องในข้อสอบวิชาภาษาอังกฤษด้วยคอมพิวเตอร์ไม่ยาก | 1 | 2 | 3 | 4 | 5 |
| 4. นักศึกษาคิดว่าเนื้อเรื่องในข้อสอบวิชาภาษาอังกฤษแบบประเพณีนิยมยาก | 1 | 2 | 3 | 4 | 5 |
| 5. นักศึกษาคิดว่าข้อสอบแบบประเพณีนิยมง่าย (กระดาษ) | 1 | 2 | 3 | 4 | 5 |
| 6. นักศึกษาคิดว่าข้อสอบด้วยคอมพิวเตอร์ยาก | 1 | 2 | 3 | 4 | 5 |
| 7. นักศึกษาคิดว่าเนื้อเรื่องที่ใช้สอบแบบประเพณีนิยม(กระดาษ) และด้วยคอมพิวเตอร์คล้ายกับเนื้อเรื่องที่เคยเรียน | 1 | 2 | 3 | 4 | 5 |
| 8. นักศึกษาชอบข้อสอบวิชาภาษาอังกฤษด้วยคอมพิวเตอร์เพราะมีเนื้อเรื่องและข้อสอบในหน้าเดียวกัน | 1 | 2 | 3 | 4 | 5 |
| 9. นักศึกษาชอบที่จะเห็นข้อสอบหลายข้อในหน้าเดียวกัน | 1 | 2 | 3 | 4 | 5 |
| 10. นักศึกษามีปัญหาในการอ่านบนจอคอมพิวเตอร์เพราะขนาดตัวหนังสือ | 1 | 2 | 3 | 4 | 5 |
| 11. นักศึกษาชอบข้อสอบแบบประเพณีนิยม(กระดาษ) เพราะเห็นเนื้อเรื่องทั้งหมดในหน้าเดียว | 1 | 2 | 3 | 4 | 5 |
| 12. นักศึกษาชอบข้อสอบวิชาภาษาอังกฤษแบบประเพณีนิยม(กระดาษ) เพราะสามารถจดบันทึกได้ | 1 | 2 | 3 | 4 | 5 |
| 13. นักศึกษาชอบข้อสอบวิชาภาษาอังกฤษด้วยคอมพิวเตอร์ เพราะสามารถทำให้เคลื่อนไหวได้ | 1 | 2 | 3 | 4 | 5 |
| 14. นักศึกษาใช้ทักษะการอ่านภาษาอังกฤษในการทำข้อสอบทั้งแบบประเพณีนิยม (กระดาษ)และด้วยคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |
| 15. นักศึกษาต้องใช้ความพยายามเพิ่มขึ้นในการทำข้อสอบด้วยคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |
| 16. นักศึกษาคิดว่าแบบฝึกหัดการใช้คอมพิวเตอร์และทำข้อสอบวิชาภาษาอังกฤษด้วยคอมพิวเตอร์ก่อนการทำข้อสอบจริงมีประโยชน์ | 1 | 2 | 3 | 4 | 5 |
| 17. นักศึกษาคิดว่าเวลาในการทำข้อสอบ (1.30 ชั่วโมง) เหมาะสม | 1 | 2 | 3 | 4 | 5 |

ตอนที่ 4 กรุณาเขียนแสดงความคิดเห็นเกี่ยวกับข้อสอบภาษาอังกฤษแบบประเพณีนิยมและแบบคอมพิวเตอร์ข้างล่าง

1. นักศึกษาต้องการที่จะทำข้อสอบวิชาภาษาอังกฤษด้วยแบบใด

คอมพิวเตอร์ เพราะ

.....

แบบประเพณีนิยม (กระดาษ) เพราะ

.....

☺A☺A ขอขอบคุณ นักศึกษาที่กรุณาตอบแบบสอบถามนี้ A☺A☺

สถาบันวิทยบริการ
 จุฬาลงกรณ์มหาวิทยาลัย

Appendix F

Frequency percentages of the students' responses to the pilot questionnaire

Part 1: My thought about the English language tests in general (N=35)

Label value: 1 strongly disagree, 2 disagree, 3 neither agree nor disagree, 4 agree, 5 strongly agree

| | | 1 | 2 | 3 | 4 | 5 |
|----|--|------|------|------|------|------|
| 1. | I do not worry about doing English tests. | 14.7 | 58.8 | 5.9 | 20.6 | - |
| 2. | When I do English tests, I often think how difficult the tests are. | 2.9 | 17.6 | 20.6 | 44.1 | 14.7 |
| 3. | Many unknown words in the reading passages discourage me to do the test. | 8.8 | 26.5 | 17.6 | 44.1 | 2.9 |
| 4. | Reading passages are interesting. | 2.9 | 8.8 | 50.0 | 29.4 | 8.8 |
| 5. | I am able to use reading skills that I have learned to do the reading tests. | - | 2.9 | 14.7 | 64.7 | 17.6 |
| 6. | When I am taking a test, I focus on the test. | - | 5.9 | 2.9 | 64.7 | 26.5 |
| 7. | I am well prepared for the tests. | - | 17.6 | 29.4 | 52.9 | - |
| 8. | I do not understand test questions. | 14.7 | 55.9 | 11.8 | 5.9 | 11.8 |
| 9. | I understand the passages. | - | 8.8 | 32.4 | 55.9 | 2.9 |

Part 2: My experience with computers (N=35)

Label value: 1 never or cannot do at all, 2 less than 3 times a week or can do not so well, 3 3-4 times a week or can do, 4 almost every day or can do well, 5 every day or can do very well

| | | 1 | 2 | 3 | 4 | 5 |
|----|--|------|------|------|------|------|
| 1. | I use the computer at home. | 2.9 | 8.8 | 20.6 | 52.9 | 14.7 |
| 2. | I use the computer at the computer lab on campus. | 5.9 | 38.2 | 23.5 | 26.5 | 5.9 |
| 3. | I use the computer at Internet cafes. | 17.6 | 58.8 | 11.8 | 8.8 | 2.9 |
| 4. | I use the computer to type reports in Thai. | 2.9 | 11.8 | 29.4 | 29.4 | 26.5 |
| 5. | I use the computer to type reports in English. | 8.8 | 29.4 | 14.7 | 35.3 | 11.8 |
| 6. | I use the computer to create graphs, draw pictures, and make presentation. | 14.7 | 29.4 | 38.2 | 11.8 | 2.9 |
| | Missing | 2.9 | | | | |
| 7. | I can use a 'mouse' or 'touch pad'. | - | 2.9 | 17.6 | 29.4 | 50.0 |
| 8. | I use the computer to surf the Internet. | - | 2.9 | 5.9 | 41.2 | 50.0 |
| 9. | I use the computer to search for academic information from websites. | - | 8.8 | 17.6 | 41.2 | 32.4 |

Part 3: My attitudes towards the RCTE from the conventional and the computer modes of assessment

Pre-test (N=35) Label value: 1 strongly disagree, 2 disagree, 3 neither agree nor disagree, 4 agree, 5 strongly agree

| | | 1 | 2 | 3 | 4 | 5 |
|-----|--|-----|------|------|------|------|
| 1. | I feel comfortable when doing the English tests on the paper-pencil mode. | 5.9 | 11.8 | 44.1 | 32.1 | 5.9 |
| 2. | I feel like doing the computer-based tests. | 2.9 | 5.9 | 26.5 | 52.9 | 11.8 |
| 3. | I think reading passages in the computer-based test are not difficult. | -- | 32.4 | 50.0 | 14.7 | -- |
| | Missing | 2.9 | | | | |
| 4. | I think reading passages in the paper-pencil test are difficult. | 5.9 | 17.6 | 38.2 | 35.3 | 2.9 |
| 5. | I think the paper-pencil tests are easy. | 2.9 | 44.1 | 47.1 | 2.9 | 2.9 |
| 6. | I think the computer-based tests are difficult. | 5.9 | 17.6 | 61.8 | 14.7 | -- |
| 7. | I think the passages in the paper-pencil and the computer-based tests are similar to those learned in class. | 5.9 | 5.9 | 52.9 | 35.3 | -- |
| 8. | I like the computer-based English test because the passage and test items are on the same screen. | -- | 5.9 | 26.5 | 47.1 | 17.6 |
| 9. | I like to see many test items on one page. | 2.9 | 14.7 | 11.8 | 52.9 | 17.6 |
| 10. | I have difficulty in reading from a screen because of font size. | 8.8 | 38.2 | 26.5 | 23.5 | 2.9 |
| 11. | I like the paper-pencil test because the reading passage is on one page. | 5.9 | 11.8 | 32.4 | 38.2 | 11.8 |
| 12. | I like the paper-pencil test because I make take notes. | | 5.9 | 29.4 | 41.2 | 23.5 |

Part 3: Pre-test Cont'd.

| | | 1 | 2 | 3 | 4 | 5 |
|-----|--|-----|------|------|------|------|
| 13. | I like doing the computer-based because of I can make some movements. | 5.9 | 17.6 | 23.5 | 41.2 | 11.8 |
| 14. | I use English reading skills when I do the tests from the paper-pencil and the computer modes. | -- | 2.9 | 17.6 | 52.9 | 26.5 |
| 15. | I have to put more effort when doing the computer-based test. | 2.9 | 20.6 | 35.3 | 26.5 | 14.7 |
| 16. | I think the tutorials priory to doing the computer-based test is useful. | -- | -- | 23.5 | 55.9 | 20.6 |
| 17. | I think the time allotment of 1.30 hours per each mode is sufficient. | 5.9 | -- | 35.3 | 55.9 | 2.9 |

Part 3: Post-test (N=35)

Label value: 1 strongly disagree, 2 disagree, 3 neither agree nor disagree, 4 agree, 5 strongly agree

| | | 1 | 2 | 3 | 4 | 5 |
|-----|--|------|------|------|------|------|
| 1. | I feel comfortable when doing the English tests on the paper-pencil mode. | 8.8 | 29.4 | 17.6 | 32.4 | 8.8 |
| 2. | I feel like doing the computer-based tests. | 2.9 | 8.8 | 32.4 | 44.1 | 8.8 |
| 3. | I think reading passages in the computer-based test are not difficult. | 2.9 | 35.3 | 32.4 | 26.5 | -- |
| | Missing | 2.9 | | | | |
| 4. | I think reading passages in the paper-pencil test are difficult. | 2.9 | 14.7 | 26.5 | 41.2 | 11.8 |
| | Missing | 2.9 | | | | |
| 5. | I think the paper-pencil tests are easy. | 14.7 | 35.3 | 35.3 | 11.8 | -- |
| | Missing | 2.9 | | | | |
| 6. | I think the computer-based tests are difficult. | 2.9 | 17.6 | 32.4 | 38.2 | 2.9 |
| | Missing | 2.9 | | | | |
| 7. | I think the passages in the paper-pencil and the computer-based tests are similar to those learned in class. | 5.9 | 17.6 | 20.6 | 47.1 | 5.9 |
| | Missing | 2.9 | | | | |
| 8. | I like the computer-based English test because the passage and test items are on the same screen. | 2.9 | 14.7 | 5.9 | 52.9 | 20.6 |
| | Missing | 2.9 | | | | |
| 9. | I like to see many test items on one page. | 2.9 | 14.7 | 11.8 | 44.1 | 23.5 |
| | Missing | 2.9 | | | | |
| 10. | I have difficulty in reading from a screen because of font size. | 14.7 | 29.4 | 29.4 | 23.5 | -- |
| | Missing | 2.9 | | | | |
| 11. | I like the paper-pencil test because the reading passage is on one page. | 8.8 | 17.6 | 26.5 | 32.4 | 11.8 |
| | Missing | 2.9 | | | | |
| 12. | I like the paper-pencil test because I make take notes. | 2.9 | 20.6 | 26.5 | 32.4 | 14.7 |
| | Missing | 2.9 | | | | |
| 13. | I like doing the computer-based because of I can make some movements. | 2.9 | 17.6 | 32.4 | 32.4 | 11.8 |
| | Missing | 2.9 | | | | |
| 14. | I use English reading skills when I do the tests from the paper-pencil and the computer modes. | 2.9 | 2.9 | 17.6 | 47.1 | 26.5 |
| | Missing | 2.9 | | | | |
| 15. | I have to put more effort when doing the computer-based test. | 14.7 | 23.5 | 29.4 | 26.5 | 2.9 |
| | Missing | 2.9 | | | | |
| 16. | I think the tutorials priory to doing the computer-based test is useful. | -- | 8.8 | 17.6 | 55.9 | 14.7 |
| | Missing | 2.9 | | | | |
| 17. | I think the time allotment of 1.30 hours per each mode is sufficient. | -- | 2.9 | 11.8 | 64.7 | 17.6 |
| | Missing | 2.9 | | | | |

Appendix G

The Attitude Questionnaire (Main Study)

แบบสอบถามทัศนคติต่อการทดสอบสัมฤทธิ์ผลการอ่านภาษาอังกฤษด้วยคอมพิวเตอร์
และแบบประเพณีนิยม (กระดาษ)

ข้อมูลของผู้ตอบแบบสอบถาม

รหัสนักศึกษา คะแนนสอบเอ็นทรานซ์/โควต้า วิชาภาษาอังกฤษ

คะแนน

Course 103 104 203 Section อายุ ปี เพศ หญิง ชาย

คำชี้แจง

แบบสอบถามนี้แบ่งออกเป็น 3 ตอน มีทั้งหมด 4 หน้า ให้นักศึกษาอ่านข้อความเกี่ยวกับการทดสอบสัมฤทธิ์ผลการอ่านภาษาอังกฤษด้วยคอมพิวเตอร์และแบบประเพณีนิยมในแต่ละส่วน แล้วเลือกกาบาทหมายเลขที่ตรงกับความคิดเห็นของนักศึกษาที่มีต่อข้อความนั้นๆ ในการตอบแบบสอบถามนี้ไม่มีการตีความคิดเห็นใดว่าถูกหรือผิด ขอความกรุณานักศึกษาตอบแบบสอบถามทุกข้อด้วย ข้อมูลที่ได้จากแบบสอบถามนี้จะนำไปใช้ในการพัฒนาการทดสอบด้านการใช้ภาษาอังกฤษด้วยคอมพิวเตอร์

ทั้งนี้ ผู้ออกแบบสอบถามขอขอบคุณนักศึกษาทุกคนที่ช่วยตอบแบบสอบถาม และขอให้สัญญาว่าข้อมูลส่วนบุคคลต่างๆที่ปรากฏในแบบสอบถามนี้จะไม่ถูกเปิดเผยแต่อย่างใด

ตัวอย่างการตอบแบบสอบถาม

ให้นักศึกษาอ่านข้อความแล้วเลือกกาบาทหมายเลขที่ตรงกับระดับความคิดเห็นของนักศึกษาที่มีต่อข้อความนั้นๆ โดยที่

- | | | |
|-----------|---------------|--|
| หมายเลข 1 | มีความหมายว่า | นักศึกษามีความเห็นด้วยน้อยหรือไม่เห็นด้วยมาก |
| หมายเลข 2 | มีความหมายว่า | นักศึกษาไม่เห็นด้วย |
| หมายเลข 3 | มีความหมายว่า | นักศึกษาไม่มีความคิดเห็นในประเด็นนี้ |
| หมายเลข 4 | มีความหมายว่า | นักศึกษามีความเห็นด้วย |
| หมายเลข 5 | มีความหมายว่า | นักศึกษามีความเห็นด้วยมาก |

| | | | | | |
|---|-------------------------------------|---|---|---|---|
| ก. ในอนาคต คอมพิวเตอร์จะมาแทนที่ครูผู้สอน | <input checked="" type="checkbox"/> | 2 | 3 | 4 | 5 |
|---|-------------------------------------|---|---|---|---|

หากนักศึกษาต้องการเปลี่ยน ให้นักศึกษาขีดทับตัวเลือกเดิม แล้วกาบาทหมายเลขใหม่ ดังนี้

| | | | | | |
|---|-------------------------------------|---|---|-------------------------------------|---|
| ก. ในอนาคต คอมพิวเตอร์จะมาแทนที่ครูผู้สอน | <input checked="" type="checkbox"/> | 2 | 3 | <input checked="" type="checkbox"/> | 5 |
|---|-------------------------------------|---|---|-------------------------------------|---|

ตอนที่ 1 ประสิทธิภาพการใช้คอมพิวเตอร์

หมายเลข 1----ไม่เคยใช้เลย หรือใช้ไม่เป็นเลย

หมายเลข 2 ---- ใช้น้อยกว่า 3 ครั้งต่อสัปดาห์

หมายเลข 3 ---- ใช้ ครั้ง 3-4 ต่อสัปดาห์ หรือ พอใช้ได้

หรือ ใช้ได้บ้าง

หมายเลข 5 -----ใช้ทุกวัน หรือ ใช้ได้ดีมาก

หมายเลข 4 -----ใช้เกือบทุกวัน หรือ ใช้ได้ดี

| | | | | | |
|---|---|---|---|---|---|
| 1. นักศึกษาใช้คอมพิวเตอร์ที่บ้าน | 1 | 2 | 3 | 4 | 5 |
| 2. นักศึกษาใช้คอมพิวเตอร์ที่มหาวิทยาลัย | 1 | 2 | 3 | 4 | 5 |
| 3. นักศึกษาใช้คอมพิวเตอร์ที่ร้านอินเทอร์เน็ต | 1 | 2 | 3 | 4 | 5 |
| 4. นักศึกษาใช้คอมพิวเตอร์พิมพ์รายงานเป็นภาษาไทย | 1 | 2 | 3 | 4 | 5 |
| 5. นักศึกษาใช้คอมพิวเตอร์พิมพ์รายงานเป็นภาษาอังกฤษ | 1 | 2 | 3 | 4 | 5 |
| 6. นักศึกษาใช้คอมพิวเตอร์สร้างกราฟ วาดรูป สร้างงานนำเสนอ | 1 | 2 | 3 | 4 | 5 |
| 7. นักศึกษาใช้ “mouse” หรือ “touch pad” | 1 | 2 | 3 | 4 | 5 |
| 8. นักศึกษาใช้คอมพิวเตอร์เพื่อท่องอินเทอร์เน็ต | 1 | 2 | 3 | 4 | 5 |
| 9. นักศึกษาใช้คอมพิวเตอร์เพื่อค้นหาข้อมูลจากเว็บไซต์ทางการศึกษา | 1 | 2 | 3 | 4 | 5 |

ตอนที่ 2 ทักษะคิดเกี่ยวกับข้อสอบภาษาอังกฤษแบบประเพณีนิยม (กระดาษ) และด้วยคอมพิวเตอร์

หมายเลข 1 – เห็นด้วยน้อยที่สุด

หมายเลข 2 -- เห็นด้วยน้อย

หมายเลข 3 – เห็นด้วยบ้าง

หมายเลข 4 – เห็นด้วยมาก

หมายเลข 5 – เห็นด้วยมากที่สุด

| | | | | | |
|--|---|---|---|---|---|
| 1. นักศึกษาเคยทำข้อสอบด้วยคอมพิวเตอร์ เช่น บน Internet หรือ website ต่างๆ | 1 | 2 | 3 | 4 | 5 |
| 2. นักศึกษารู้สึกวิตกกังวลเมื่อจะทำข้อสอบ | | | | | |
| 2.1 แบบประเพณีนิยม (บนกระดาษ) | 1 | 2 | 3 | 4 | 5 |
| 2.2 ด้วยคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |
| 3. นักศึกษาคิดว่าข้อสอบแบบปรนัย (เลือกตอบ) เหมาะสม กับการทดสอบ | | | | | |
| 3.1 แบบประเพณีนิยม (บนกระดาษ) | 1 | 2 | 3 | 4 | 5 |
| 3.2 ด้วยคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |
| 4. นักศึกษาคิดว่าความยาวของเนื้อเรื่อง (ประมาณ 350 – 400 คำ หรือ เท่ากับกระดาษ A 4) มีความเหมาะสมกับข้อสอบ | | | | | |
| 4.1 แบบประเพณีนิยม (บนกระดาษ) | 1 | 2 | 3 | 4 | 5 |
| 4.2 ด้วยคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |
| 5. นักศึกษาคิดว่าข้อสอบการอ่านภาษาอังกฤษ | | | | | |
| 5.1 แบบประเพณีนิยม (บนกระดาษ) มีความยาก | 1 | 2 | 3 | 4 | 5 |
| 5.2 ด้วยคอมพิวเตอร์ มีความยาก | 1 | 2 | 3 | 4 | 5 |

| | | | | | |
|---|---|---|---|---|---|
| 6. นักศึกษาชอบคำสั่งในการทำข้อสอบที่เป็นภาษาไทย | 1 | 2 | 3 | 4 | 5 |
| 7. นักศึกษาชอบคำสั่งในการทำข้อสอบที่เป็นภาษาอังกฤษ | 1 | 2 | 3 | 4 | 5 |
| 8. นักศึกษาคิดว่าขนาดของตัวอักษรภาษาอังกฤษ Times New Roman Size 13---
ขนาดเท่ากับ TIMES นี้) เหมาะสมในการพิมพ์ข้อทดสอบ | | | | | |
| 8.1 แบบประเพณีนิยม (บนกระดาษ) | 1 | 2 | 3 | 4 | 5 |
| 8.2 ด้วยคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |
| 9. นักศึกษาชอบการทำข้อทดสอบ | | | | | |
| 9.1 แบบประเพณีนิยม (บนกระดาษ) เพราะ | | | | | |
| 9.1.1 สามารถพลิกหน้ากระดาษไปมา | 1 | 2 | 3 | 4 | 5 |
| 9.1.2 สามารถขีดเขียนทำบันทึกได้ | 1 | 2 | 3 | 4 | 5 |
| 9.1.3 มีข้อสอบทุกข้อยาวต่อเนื่องกันไป | 1 | 2 | 3 | 4 | 5 |
| 9.1.4 สามารถดูข้อสอบหลายข้อในเวลาเดียวกันได้ | 1 | 2 | 3 | 4 | 5 |
| 9.1.5 เชื่อมั่นในการตรวจข้อสอบอย่างเที่ยงตรงของอาจารย์ | 1 | 2 | 3 | 4 | 5 |
| 9.2 ด้วยคอมพิวเตอร์ เพราะ | | | | | |
| 9.2.1 ใช้การคลิกปุ่มบอกทิศทางโดยไม่ต้องพลิกหน้า
กระดาษ | 1 | 2 | 3 | 4 | 5 |
| 9.2.2 สามารถทำให้มีการเคลื่อนไหวโดยการลาก
scrolling bar | 1 | 2 | 3 | 4 | 5 |
| 9.2.3 นักศึกษารู้สึกตื่นตัวเวลาทำข้อสอบ | 1 | 2 | 3 | 4 | 5 |
| 9. นักศึกษาชอบการทำข้อทดสอบ | | | | | |
| 9.2 ด้วยคอมพิวเตอร์ เพราะ | | | | | |
| 9.2.4 มีเนื้อเรื่องอยู่ด้านใดด้านหนึ่งและมีข้อสอบอยู่ด้านข้าง
ทีละข้อ อาทิเช่น บนหน้าจอคอมพิวเตอร์เดียวกัน | 1 | 2 | 3 | 4 | 5 |
| 9.2.5 เชื่อมั่นในการคำนวณของคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |
| 9.2.6 สามารถรู้ผลข้อสอบได้ทันที | 1 | 2 | 3 | 4 | 5 |
| 10. นักศึกษาคิดว่าในเวลา 1.00 ชั่วโมงนักศึกษามีเวลาเพียงพอสำหรับทำข้อ
สอบ | | | | | |
| 10.1 แบบประเพณีนิยม (บนกระดาษ) | 1 | 2 | 3 | 4 | 5 |
| 10.2 ด้วยคอมพิวเตอร์ | 1 | 2 | 3 | 4 | 5 |

ตอนที่ 3 กรุณาเขียนแสดงความคิดเห็นเกี่ยวกับข้อสอบภาษาอังกฤษแบบประเพณีนิยมและแบบคอมพิวเตอร์ข้างล่าง

1. ความคิดเห็นของนักศึกษาต่อการทำข้อสอบวิชาภาษาอังกฤษ:

| | |
|--|---|
| <p>ชอบทำข้อสอบแบบประเพณีนิยม(กระดาษ)
เพราะ.....
.....
.....
.....
.....</p> | <p>ไม่ชอบทำข้อสอบแบบประเพณีนิยม (กระดาษ) เพราะ
.....
.....
.....
.....
.....</p> |
| <p>ชอบทำข้อสอบด้วยคอมพิวเตอร์ เพราะ
.....
.....
.....
.....
.....</p> | <p>ไม่ชอบทำข้อสอบด้วยคอมพิวเตอร์ เพราะ
.....
.....
.....
.....
.....</p> |

☺A☺A

ขอขอบคุณ นักศึกษาที่กรุณาตอบแบบสอบถามนี้

A☺A☺

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix H

Frequency percentages of the students' responses to the main study questionnaire

Part 1: My experience with computers (N=66)

Label value: 1 never or cannot do at all, 2 less than 3 times a week or can do not so well, 3 3-4 times a week or can do, 4 almost every day or can do well, 5 every day or can do very well

| | 1 | 2 | 3 | 4 | 5 |
|---|------|------|------|------|------|
| 1. I use the computer at home. | 9.1 | 19.7 | 19.7 | 34.8 | 16.7 |
| 2. I use the computer at the computer lab on campus. | 1.5 | 24.2 | 34.8 | 22.7 | 16.7 |
| 3. I use the computer at Internet cafes. | 10.6 | 63.6 | 15.2 | 4.5 | 6.1 |
| 4. I use the computer to type reports in Thai. | 3.0 | 18.2 | 24.2 | 34.8 | 19.7 |
| 5. I use the computer to type reports in English. | 7.6 | 25.8 | 34.8 | 21.2 | 10.6 |
| 6. I use the computer to create graphs, draw pictures, and make presentation. | 13.6 | 47.0 | 19.7 | 16.7 | 3.0 |
| 7. I can use a 'mouse' or 'touch pad'. | 4.5 | 7.6 | 24.2 | 40.9 | 22.7 |
| 8. I use the computer to surf the Internet. | 1.5 | 3.0 | 16.7 | 40.9 | 37.9 |
| 9. I use the computer to search for academic information from websites. | 1.5 | 6.1 | 21.2 | 43.9 | 27.3 |

Part 2: My attitudes towards the Reading Comprehension Test of English (RCTE) from the conventional and the computer modes of assessment (N=66)

Pre-test

Label value: 1 strongly disagree, 2 disagree, 3 neither agree nor disagree, 4 agree, 5 strongly agree

| | 1 | 2 | 3 | 4 | 5 |
|--|------|------|------|------|------|
| 1. I have done computer-based tests available on the Internet and websites. | 28.8 | 18.2 | 22.7 | 16.7 | 3.0 |
| Missing | 10.6 | | | | |
| 2. I am worried when doing | | | | | |
| 2.1 the paper-based test | 12.1 | 4.5 | 39.4 | 30.3 | 9.1 |
| Missing | 4.5 | | | | |
| 2.2 the computer-based test | 3.0 | 13.6 | 54.5 | 18.2 | 9.1 |
| Missing | 1.5 | | | | |
| 3. I think the multiple-choice test format is appropriate for | | | | | |
| 3.1 the paper-based test | 3.0 | 7.6 | 21.2 | 43.9 | 21.2 |
| Missing | 3.0 | | | | |
| 3.2 the computer-based test | 3.0 | 3.0 | 27.3 | 39.4 | 21.2 |
| Missing | 6.1 | | | | |
| 4. I think a reading passage with the length of 350-400 or on an A-4 page is appropriate for | | | | | |
| 4.1 the paper-based test | 3.0 | 24.2 | 31.8 | 31.8 | 9.1 |
| 4.2 the computer-based test | 13.6 | 28.8 | 33.3 | 18.2 | -- |
| Missing | 6.1 | | | | |
| 5. I think the English reading tests on | | | | | |
| 5.1 the paper is difficult. | 3.0 | 10.6 | 36.4 | 37.9 | 10.6 |
| Missing | 1.5 | | | | |
| 5.2 the computer is difficult. | -- | 6.1 | 40.9 | 37.9 | 13.6 |
| Missing | 1.5 | | | | |
| 6. I like the instructions in Thai. | -- | 12.1 | 36.4 | 27.3 | 24.2 |
| 7. I like the instructions in English. | 6.1 | 24.2 | 43.9 | 22.7 | 3.0 |
| 8. I think that the font size of Times New Roman 13 (TIMES) is appropriate for | | | | | |
| 8.1 the paper-based test | 4.5 | 13.6 | 36.4 | 33.3 | 7.6 |
| Missing | 4.5 | | | | |
| 8.2 the computer-based test | 4.5 | 10.6 | 27.3 | 42.4 | 13.6 |
| Missing | 1.5 | | | | |
| 9. I like doing the English test on | | | | | |

| | | | | | | |
|-----|--|------|------|------|------|------|
| | 9.1 the paper because | | | | | |
| | 9.1.1 I can touch and turn pages. | 10.6 | 16.7 | 28.8 | 27.3 | 16.7 |
| | 9.1.2 I can take/make notes. | -- | 4.5 | 12.1 | 54.5 | 28.8 |
| | 9.1.3 I can see list of test items. | 1.5 | 10.6 | 42.4 | 34.8 | 9.1 |
| | Missing | | | 1.5 | | |
| | 9.1.4 I can view many test items at once. | 3.0 | 4.5 | 19.7 | 56.1 | 16.7 |
| | 9.1.5 I trust the marking of the teacher. | 6.1 | 18.2 | 51.5 | 16.7 | 4.5 |
| | Missing | | | 3.0 | | |
| | 9.2 the computer because | | | | | |
| | 9.2.1 I can click navigate icons to turn pages. | 4.5 | 7.6 | 34.8 | 37.9 | 13.6 |
| | Missing | | | 1.5 | | |
| | 9.2.2 I can make movements by scrolling. | 3.0 | 7.6 | 48.5 | 25.8 | 13.6 |
| | Missing | | | 1.5 | | |
| | 9.2.3 I feel being active while doing it. | 7.6 | 13.6 | 45.5 | 24.2 | 7.6 |
| | Missing | | | 1.5 | | |
| | 9.2.4 I can see the passage on one side of a screen and a test item on the other side. | 6.1 | 6.1 | 39.4 | 40.9 | 4.5 |
| | Missing | | | 3.0 | | |
| | 9.2.5 I trust the marking of the computer. | 3.0 | 4.5 | 39.4 | 37.9 | 15.2 |
| | 9.2.6 I can receive score report immediately when finished. | 1.5 | -- | 9.1 | 50.0 | 37.9 |
| | Missing | | | 1.5 | | |
| 10. | I think that the time allotment of 1 hour is sufficient for | | | | | |
| | 10.1 the paper-based test | 4.5 | 28.8 | 42.4 | 18.2 | 3.0 |
| | Missing | | | 3.0 | | |
| | 10.2 the computer-based test | 4.5 | 22.7 | 39.4 | 30.3 | 3.0 |

Part 2: Post-test

Label value: 1 strongly disagree, 2 disagree, 3 neither agree nor disagree, 4 agree, 5 strongly agree

| | | | | | | |
|----|---|------|------|------|------|------|
| | | 1 | 2 | 3 | 4 | 5 |
| 1. | I have done computer-based tests available on the Internet and websites. | 13.6 | 18.2 | 30.3 | 9.1 | 7.6 |
| | Missing | | | 21.2 | | |
| 2. | I am worried when doing | | | | | |
| | 2.1 the paper-based test | 3.0 | 10.6 | 51.5 | 19.7 | 12.1 |
| | Missing | | | 3.0 | | |
| | 2.2 the computer-based test | 1.5 | 7.6 | 53.0 | 16.7 | 15.2 |
| | Missing | | | 6.1 | | |
| 3. | I think the multiple-choice test format is appropriate for | | | | | |
| | 3.1 the paper-based test | 4.5 | 7.6 | 28.8 | 31.8 | 24.2 |
| | Missing | | | 3.0 | | |
| | 3.2 the computer-based test | 4.5 | 10.6 | 28.8 | 37.9 | 16.7 |
| | Missing | | | 1.5 | | |
| 4. | I think a reading passage with the length of 350-400 or on an A-4 page is appropriate for | | | | | |
| | 4.1 the paper-based test | 4.5 | 18.2 | 34.8 | 34.8 | 6.1 |
| | Missing | | | 1.5 | | |
| | 4.2 the computer-based test | 9.1 | 15.2 | 48.5 | 18.2 | 6.1 |
| | Missing | | | 3.0 | | |
| 5. | I think the English reading tests on | | | | | |
| | 5.1 the paper is difficult. | 4.5 | 4.5 | 25.8 | 48.5 | 6.7 |
| | 5.2 the computer is difficult. | 3.0 | 12.1 | 47.0 | 22.7 | 12.1 |
| | Missing | | | 3.0 | | |
| 6. | I like the instructions in Thai. | 1.5 | 9.1 | 25.8 | 33.3 | 28.8 |
| | Missing | | | 1.5 | | |
| 7. | I like the instructions in English. | 3.0 | 18.2 | 53.0 | 22.7 | 1.5 |
| | Missing | | | 1.5 | | |

| | | | | | | |
|---------|--|------|------|------|------|------|
| 8. | I think that the font size of Times New Roman 13 (TIMES) is appropriate for | | | | | |
| | 8.1 the paper-based test | 3.0 | 12.1 | 47.0 | 22.7 | 12.1 |
| | Missing | 3.0 | | | | |
| | 8.2 the computer-based test | 3.0 | 6.1 | 36.4 | 36.4 | 16.7 |
| | Missing | 1.5 | | | | |
| 9. | I like doing the English test on | | | | | |
| | 9.1 the paper because | | | | | |
| | 9.1.1 I can touch and turn pages. | 12.1 | 9.1 | 34.8 | 28.8 | 15.2 |
| | 9.1.2 I can take/make notes. | 1.5 | 3.0 | 18.2 | 50.0 | 27.3 |
| | 9.1.3 I can see list of test items. | 3.0 | 12.1 | 42.4 | 30.3 | 12.1 |
| | 9.1.4 I can view many test items at once. | 3.0 | 6.1 | 31.8 | 42.4 | 16.7 |
| | 9.1.5 I trust the marking of the teacher. | 4.5 | 18.2 | 62.1 | 12.1 | 3.0 |
| | 9.2 the computer because | | | | | |
| | 9.2.1 I can click navigate icons to turn pages. | 3.0 | 4.5 | 22.7 | 50.0 | 19.7 |
| | 9.2.2 I can make movements by scrolling. | 3.0 | 4.5 | 34.8 | 40.9 | 16.7 |
| | 9.2.3 I feel being active while doing it. | 3.0 | 10.6 | 51.5 | 21.2 | 13.6 |
| | 9.2.4 I can see the passage on one side of a screen and a test item on the other side. | 1.5 | 3.0 | 39.4 | 33.3 | 21.2 |
| | Missing | 1.5 | | | | |
| | 9.2.5 I trust the marking of the computer. | 3.0 | 1.5 | 21.2 | 53.0 | 19.7 |
| | Missing | 1.5 | | | | |
| | 9.2.6 I can receive score report immediately when finished. | -- | 1.5 | 7.6 | 45.5 | 42.4 |
| Missing | 3.0 | | | | | |
| 10. | I think that the time allotment of 1 hour is sufficient for | | | | | |
| | 10.1 the paper-based test | 9.1 | 16.7 | 45.5 | 22.7 | 4.5 |
| | Missing | 1.5 | | | | |
| | 10.2 the computer-based test | 7.6 | 12.1 | 48.5 | 24.2 | 6.1 |
| | Missing | 1.5 | | | | |

Mean Scores of the Attitude Questionnaire (Part 2) before and after the Test Administration from the Main Study

| | | Pre-test \bar{X} | Post-test \bar{X} |
|-----|---|--------------------|---------------------|
| 1. | I have done computer-based tests available on the Internet and websites. | 2.15 | 2.15 |
| 2. | I am worried when doing | | |
| | 2.1 the paper-based test | 3.06 | 3.18 |
| | 2.2 the computer-based test | 3.12 | 3.18 |
| 3. | I think the multiple-choice test format is appropriate for | | |
| | 3.1 the paper-based test | 3.63 | 3.54 |
| | 3.2 the computer-based test | 3.54 | 3.46 |
| 4. | I think a reading passage with the length of 350-400 or on an A-4 page is appropriate for | | |
| | 4.1 the paper-based test | 3.19 | 3.15 |
| | 4.2 the computer-based test | 2.43 | 2.87 |
| 5. | I think the English reading tests on | | |
| | 5.1 the paper is difficult. | 3.37 | 3.68 |
| | 5.2 the computer is difficult. | 3.54 | 3.54 |
| 6. | I like the instructions in Thai. | 3.63 | 3.74 |
| 7. | I like the instructions in English. | 2.92 | 2.96 |
| 8. | I think that the font size of Times New Roman 13 (TIMES) is appropriate for | | |
| | 8.1 the paper-based test | 3.12 | 3.19 |
| | 8.2 the computer-based test | 3.45 | 3.53 |
| 9. | I like doing the English test on | | |
| | 9.1 the paper because | | |
| | 9.1.1 I can touch and turn pages. | 3.22 | 3.25 |
| | 9.1.2 I can take/make notes. | 4.07 | 3.98 |
| | 9.1.3 I can see list of test items. | 3.34 | 3.36 |
| | 9.1.4 I can view many test items at once. | 3.78 | 3.63 |
| | 9.1.5 I trust the marking of the teacher. | 2.86 | 2.90 |
| | 9.2 the computer because | | |
| | 9.2.1 I can click navigate icons to turn pages. | 3.43 | 3.78 |
| | 9.2.2 I can make movements by scrolling. | 3.34 | 3.63 |
| | 9.2.3 I feel being active while doing it. | 3.06 | 3.31 |
| | 9.2.4 I can see the passage on one side of a screen and a test item on the other side. | 3.22 | 3.65 |
| | 9.2.5 I trust the marking of the computer. | 3.57 | 3.80 |
| | 9.2.6 I can receive score report immediately when finished. | 4.18 | 4.19 |
| 10. | I think that the time allotment of 1 hour is sufficient for | | |
| | 10.1 the paper-based test | 2.77 | 2.92 |
| | 10.2 the computer-based test | 3.04 | 3.04 |

BIOGRAPHY

Miss Sudthida Pavavijarn was born on July 23, 1975 in Ubonratchatani, Thailand. She received a Bachelor's Degree of Arts in English (Honors) from Chiang Mai University in the year 2000. She was granted a scholarship for developing teaching staff of Chiang Mai University while she was in her senior year. Therefore, immediately after graduation she became an English language lecturer at the Department of English, Faculty of Humanities, Chiang Mai University. In the year 2002 she was admitted to study in a postgraduate program in English as an International Language at Chulalongkorn University under a scholarship for developing teaching staff of Chiang Mai University. At present, she is working as an English language lecturer at the Department of English, Faculty of Humanities, Chiang Mai University.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย