การใช้แบบทดสอบซีบนระบบเครือข่ายเพื่อจัดระดับความสามารถด้านภาษาอังกฤษ ของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก และการศึกษากลวิธีการตอบข้อสอบ

นางณัฐาภร คำมะสอน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต
สาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ (สหสาขาวิชา)
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2551

THE USE OF THE WEB-BASED C-TEST (WBCT) AS AN ENGLISH PLACEMENT TEST FOR RAJAMANGALA UNIVERSITY OF TECHNOLOGY TAWAN-OK AND AN INVESTIGATION OF TEST-TAKERS' STRATEGIES

Mrs. Naddhaporn Kammasorn

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in English as an International Language
(Interdisciplinary Program)
Graduate School
Chulalongkorn University
Academic Year 2008

Thesis Title	The Use of the Web-based C-Test (WBCT) as an English Placement Test for Rajamangala University of Technology Tawan-ok and an Investigation of Test-takers' Strategies
By	Mrs. Naddhaporn Kammasorn
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 Doctoral Degree

..... Dean of the Graduate School
(Associate Professor Pornpote Piumsomboon, Ph.D.)

THESIS COMMITTEE

..... Chairman
(Assistant Professor Kulaporn Hiranburana, Ph.D.)

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

..... External Member
(Assistant Professor Somsak Boonsathorn, Ph.D.)

..... Member
(Assistant Professor Pavinee Thirakhupt, Ph.D.)

..... Member
(Associate Professor Prakaikaew Opanon-Amata.)

ณัฐาภร คำมะสอน: การใช้แบบทดสอบซีบนระบบเครือข่ายเพื่อจัดระดับความสามารถด้านภาษาอังกฤษของนักศึกษา มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก และการศึกษากลวิธีการตอบข้อสอบ (THE USE OF THE WEB-BASED C-TEST (WBCT) AS AN ENGLISH PLACEMENT TEST FOR RAJAMANGALA UNIVERSITY OF TECHNOLOGY TAWAN-OK AND AN INVESTIGATION OF TEST-TAKERS' STRATEGIES) อ. ที่ปรึกษา: ศ. ดร. กาญจนา ปราบพาล, 165 หน้า.

การวิจัยนี้ศึกษาการใช้แบบทดสอบซีบนระบบเครือข่ายเพื่อจัดระดับความสามารถด้านภาษาอังกฤษของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก และศึกษากลวิธีการตอบข้อสอบ รวมทั้งศึกษาความคิดเห็นของนักศึกษาต่อแบบทดสอบซีบนระบบเครือข่าย กลุ่มตัวอย่างเป็นนักศึกษาชั้นปีที่ 1 จำนวน 134 คน เครื่องมือในการวิจัยเป็นแบบทดสอบ Quick Placement Test (QPT) แบบทดสอบซีบนระบบเครือข่าย (WBCT) และการสัมภาษณ์ภายหลังการทำแบบทดสอบ ในการวิเคราะห์ข้อมูล ผู้วิจัยหาความเชื่อมั่นของแบบทดสอบซีบนระบบเครือข่าย หาค่าสหสัมพันธ์ของแบบทดสอบซีบนระบบเครือข่าย (WBCT) กับแบบทดสอบ Quick Placement Test (QPT) และจัดระดับความสามารถโดยใช้แบบทดสอบซีบนระบบเครือข่าย นอกจากนี้ วิเคราะห์เชิงเนื้อหาเกี่ยวกับกลวิธีการตอบข้อสอบ และ ความคิดเห็นของนักศึกษาต่อแบบทดสอบซีบนระบบเครือข่าย

ผลการวิจัยพบว่า แบบทดสอบ WBCT มีค่าความเชื่อมั่น .817 (Cronbach's Alpha) และจากการหาค่าสหสัมพันธ์เพียร์สัน พบว่าแบบทดสอบ WBCT และแบบทดสอบ QPT มีความสัมพันธ์กันอย่างมีนัยสำคัญที่ .01 ค่าระดับความสัมพันธ์ คือ .340 แบบทดสอบ WBCT สามารถจัดระดับความสามารถด้านภาษาอังกฤษของนักศึกษาได้ 2 ระดับ ตามการเทียบแบ่งระดับของสถาบันทดสอบภาษาในยุโรป (ALTE) คือ ระดับ 0 หรือระดับผู้เริ่มเรียน และ ระดับ 1 หรือระดับปฐมภูมิ เมื่อทดสอบความแตกต่างของทั้ง 2 ระดับ โดยใช้ t - test พบว่า มีความแตกต่างอย่างมีนัยสำคัญที่ .01 ในการศึกษากลวิธีในการตอบข้อสอบพบว่า นักศึกษาที่ทำคะแนนสูง ใช้กลวิธีการทำข้อสอบโดย 1) การอ่านประโยค และ 2) การแปลคำและประโยคในบริบท ในขณะที่นักศึกษาที่ทำคะแนนต่ำใช้กลวิธีการทำข้อสอบโดย 1) การเดา และ 2) การอ่านแต่ละคำหรือส่วนของประโยค นอกจากนี้ยังพบว่า นักศึกษาที่ทำคะแนนสูงใช้ความรู้ด้านคำศัพท์และความสามารถภาษาอังกฤษด้านอื่นในการเติมคำ ในขณะที่นักศึกษาที่ทำคะแนนต่ำใช้ความรู้ด้านคำศัพท์อย่างเดียว นักศึกษาที่ทำคะแนนสูงและนักศึกษาที่ทำคะแนนต่ำใช้ประโยชน์จากอักษรที่ให้ครึ่งคำในแบบทดสอบช่วยในการตอบ ในการศึกษาด้านความคิดเห็นต่อแบบทดสอบนี้ พบว่านักศึกษาที่ได้คะแนนสูงส่วนใหญ่เชื่อว่าแบบทดสอบนี้วัดความสามารถภาษาอังกฤษทุกด้าน ส่วนนักศึกษาที่ได้คะแนนต่ำส่วนใหญ่เชื่อว่าแบบทดสอบนี้วัดความสามารถด้านคำศัพท์เท่านั้น นอกจากนี้ นักศึกษาที่ได้คะแนนสูงและนักศึกษาที่ได้คะแนนต่ำคิดว่าแบบทดสอบซีบนระบบเครือข่ายน่าสนใจ ท้าทาย และ ยาก และนักศึกษาชอบทำแบบทดสอบบนระบบเครือข่าย

สาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ (สหสาขาวิชา) ลายมือชื่อนิสิต .....

ปีการศึกษา 2551 ลายมือชื่ออาจารย์ที่ปรึกษา.....

4889660820: MAJOR ENGLISH AS AN INTERNATIONAL LANGUAGE
KEY WORDS: WEB-BASED TEST/ C-TEST/ PLACEMENT TEST/ TEST-TAKERS' STRATEGIES/ ENGLISH ASSESSEMENT

NADDHAPORN KAMMASORN: THE USE OF THE WEB-BASED C-TEST (WBCT) AS AN ENGLISH PLACEMENT TEST FOR RAJAMANGALA UNIVERSITY OF TECHNOLOGY TAWAN-OK AND AN INVESTIGATION OF TEST-TAKERS' STRATEGIES. THESIS ADVISOR: PROF. KANCHANA PRAPPHAL, Ph.D., 165 pp.

This study investigated the use of the Web-based C-Test as a placement test for students at Rajamangala University of Technology Tawan-ok. Also, proficiency levels classified by the Web-based C-Test were explored. Additionally, it investigated the test-takers' strategies and their opinions in doing the Web-based C-Test. The subjects were 134 first-year, undergraduate students. The research instruments were the Quick Placement Test (QPT) as a criterion test, the Web-based C-Test (WBCT), and a retrospective interview. The reliability of the WBCT, the correlation between the WBCT and the QPT, and the level classification by the WBCT were reported. Moreover, content analysis was employed to evaluate the test-takers' strategies and their opinions towards the test.

The findings revealed that the Web-based C-Test had a reliability of .817 (Cronbach's Alpha). Pearson correlation coefficient between the WBCT and the QPT was .340 at the .01 level. The WBCT could differentiate the students into two levels according to the Association of Language Testers in Europe (ALTE) levels; Level 0 or 'Beginner' and Level 1 or "Elementary." The two levels were significantly different at .01 calculated by a t–test. Additionally, the "high" and "low" achievers implemented different strategies in doing the Web-based C-Test. The "high" achievers employed 1) reading the sentence and 2) translating words and sentences in the context while the "low" achievers used 1) guessing and 2) reading each word or parts of sentences. In addition, the "high" achievers employed vocabulary and general proficiency in restoring the words while the "low" achievers employed only vocabulary in restoring the words. Moreover, both "high" and "low" achievers made use of the first half of the words. Furthermore, the "high" and "low" achievers expressed different opinions towards the WBCT. The majority of the "high" achievers accepted that the test tested their general proficiency while the majority of the "low" achievers expressed that the test tested vocabulary. Finally, both "high" and "low" achievers stated that the WBCT was interesting, challenging and difficult and they liked the web-based administration.

Field of study: English as an International Language Student's signature.....
(Interdisciplinary Program) Advisor's signature Kanchana Prapphal

Academic year: 2008

ACKNOWLEDGEMENTS

This research study was made possible because of the great help I received from so many. I wish to express my gratitude to all of them. First and foremost, I am indebted to the three professors: Professor Dr. Kanchana Prapphal, my supervisor, to whom I am grateful for her devoted time and invaluable guidance throughout the study; Professor Dr. Ulrich Raatz, one of the initiators of C-testing, who supported me with precious advice and documentation; and Joe Anthony Blum, who contributed the software program for this study and also provided me with excellent advice on editing my work. I am highly thankful to my research committee who devoted their time and patience to read my work: Assistant Professor Dr. Kulaporn Hiranburana, who chaired the dissertation committee; Associate Professor Prakaikaew Opanon-Amata; Assistant Professor Dr. Pavinee Thirakhupt; and Assistant Professor Dr. Somsak Boonsathorn, the external committee member from Mae Fah Luang University, who initially inspired me with the C-tests on his lectures at NIDA. My deepest appreciation goes to the three English language experts who committed their valuable time to rate my questionnaire for text selection: Associate Professor Antikar Rongsa-Ard, Associate Professor Dr. Supanee Chinnawongs, and Associate Professor Dr. Punchalee Wasanasomsithi. Also, special thanks go to the three English native speakers who assisted in doing the Web-based C-Test: Ajarn Andrew Mathews and Ajarn Andrew Ross from Chulalongkorn University Language Institute; and Dr. Robert Troyer, my EIL friend. During my bewildering moment dealing with the data, there was a person who patiently contributed his time to do away with the clouds. My heartfelt thank goes to Ajarn Thawisakdi Dechalert who assisted me with the data analysis. I am also indebted to my students who kindly did the test for my research study. Additionally, I wish to thank all my teachers and all the professors at the EIL program from whom I adopted the jewels of never-ending love of hard work. My wholehearted thanks also go to my friends and family, whose love and encouragements are the sources of my strengths to continue my work. Besides, I am indebted to all the shortcomings that not only pushed me ahead but also lit me the path to my destination. Finally, I am grateful to the 90th Anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund) for supporting my research study.

CONTENTS

CHAPTER PAGE

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

TABLE PAGE

LIST OF EXAMPLES

EXAMPLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

Background of the Study

As a common practice in Thailand, students entering universities are required to take fundamental English courses in their early years of studies. It is, therefore, necessary for educational institutions to have tests that can place students at appropriate levels before they enter their study program because placement tests can provide information that helps place students at their appropriate ability level in the teaching program (Hughes, 2003). Through them, students can be grouped homogeneously based on their level of language ability (Bachman, 1990). Therefore, teachers can benefit from placement decisions because their classes are comprised of students with relatively similar language skills (Brown, 1996). It is obvious that placement tests are essential in performing the function of placing students with others at the same ability level. In doing so, students with different proficiencies are placed in different levels so that all may benefit from having the type of encouraging atmosphere found only in same-ability groups.

However, some institutions are aware of the necessity of placement tests while others forego the practice. Naturally, this might be because of some constraints; for example, there may be no appropriate placement tests or the placement tests are beyond their budget. Thus, without placement testing, students are inevitably seated in English classes with respect to their major studies instead of their English proficiency -- a practice maybe appropriate for some, but devastating to others. It is essential that educational institutions should provide efficient placement tests to place students in appropriate levels at the very beginning of the study. This not only facilitates teachers in adjusting the learning environment appropriately to the students' levels, but also encourages improvements in teaching and learning appropriate to learners' abilities.

As for Rajamangala University of Technology Tawan-ok (RMUTTO), up until now, there was never a budget for commercial English placement tests. As a result,

students were grouped into English classes according to their major studies. This has caused problems in teaching and learning because of the differences in the students' proficiency. To solve the problems, it would be advisable that the institution create its own placement tests. This agrees with Brown's idea (2005: 13) stating that "a placement test developed for one institution would not be appropriate for students in other institutions." He stresses that an institution should construct its own placement tests. Poel and Weatherly (1997) go even further, saying that commercial placement tests rarely ever place learners into their appropriate levels. In this regard, it would be advisable for RMUTTO to develop its own placement tests to group students into appropriate levels. This would result in a better organization and facilitate the teaching and learning process. The prospective test should not only differentiate but also benchmark students' proficiency levels. Importantly, the test should cost the institution as little as possible for development, administration, and scoring.

In constructing a placement test, test developers should take methods of assessing learners' knowledge into consideration. Among proficiency tests, cloze tests have been reported to be valid tests (Cohen, 1980). Cohen points out that there are high correlations between cloze test scores and those of reading comprehension, writing ability, listening ability, and speaking proficiency, which indicates that the cloze test evaluates overall language proficiency. Bachman (1982) expresses that cloze tests can assess both low-level skills like phrase processing and complex skills like discourse processing. McNamara (2000) supports that a cloze test assesses productive skills, speaking and writing in particular, because readers are required to integrate their grammatical, lexical, contextual, and pragmatic knowledge to supply the missing words. According to Cohen, Bachman and McNamara, several skills can be assessed simultaneously through a cloze test, which is good in that a cloze test reveals learners' overall proficiency.

Despite the usefulness of cloze tests in language assessment, there are criticisms of cloze tests. According to Alderson's studies (1979, 1983) a cloze test has to be a long text to ensure enough items and one long text may lead to test bias; the reliability and validity of cloze tests are affected by the deletion rate and the starting point; the exact method of scoring makes the test too difficult while the acceptable method of

scoring makes the test subjective and becomes less reliable; and the test difficulty depends much on the proportion of deletions of the structure and content words. Therefore, in order to adopt the cloze test for measurement of students' abilities, particularly, to serve a certain purpose of the institution, these deficiencies must be taken into account before making decisions whether to use it or modify it. As a result, to solve the problems of cloze tests, there was an attempt to modify cloze tests by Raatz and Klein-Braley (1981) who developed C-tests as an improvement of cloze tests. The two tests have similarities and differences in various aspects.

In terms of similarities, C-tests are based on the same principles as cloze tests in that they are building on the reduced redundancy theory, stating that a redundant message contains more information than is necessary to understand the message. Thus, when some parts of the message are damaged or omitted, the left message can still convey the understanding of the message. As for the differences, cloze tests and C-tests are different in text selection. A long passage is implemented in a cloze test while 5 to 6 short passages are employed in a C-test. Also, the two tests differ in the deletion techniques. In cloze tests, whole words are deleted at regular intervals while in C-tests only the second half of every other word is deleted leaving the first halves as clues for restoration. This means that the undamaged parts make a virtually full restoration possible, at least for native speakers of a language. Naturally, the deletion of the second half of every second word, which is the C-test format, makes the deletions occur on both function words and content words. As a result, the test-takers have to employ semantic, syntactic, morphological, lexical, and orthographic processing to integrate contextual information in restoring the missing parts (Hastings, 2002b). According to the C-test approach, it is not the redundancy of the text that is measured, but the test-takers' ability to employ their language ability to restore the missing parts of the text (Raatz and Klein-Braley, 1981). Additionally, cloze tests and C-tests vary in terms of scoring. The cloze tests implement either the exact method of scoring or the acceptable method of scores because in the whole blank space there is a probability that other words rather than the key word are accepted as correct while the second half of the word deleted in the C-tests accept

only the exact method of scoring. This makes the C-test more objective and it also enhances its reliability.

It is apparent that C-tests are superior in terms of deletion techniques and scoring methods. The deletion techniques in C-tests provide test-takers some clues to rely on. Also, the exact method of scoring makes C-tests more objective. However, there are some constraints in text selection, which lead to precision in selecting and editing the five or six chosen paragraphs until they meet all of the C-test criteria. The advantages of C-tests in choice of deletion techniques and using the exact method of scoring attracted the researcher to implement a C-test for this study. Importantly, Grotjahn et al (2002) reported that despite the fewer studies, C-tests have been confirmed as sophisticated language tests. Therefore, the C-test procedure was employed in this study although there has been no prior study employing the C-test as an English placement test in the Thai context.

In terms of test delivery methods, the Web-based methods have been extensively improvised in language testing in that the Web-based tests excel Paper and Pencil Tests (P&P) in many ways. According to the Department of Applied Linguistics and TESL of The University of California, Los Angeles, (UCLA, n.d), there are several important limitations of Paper & Pencil Tests; namely, personnel, scheduling, authenticity of assessment tasks, and the use of new measurement technology. In regard to personnel, the limitations occur in preparing new assessment, administering the tests, scoring student responses, and preparing reports. For scheduling, the scheduling of the test administration is not convenient for all students. Additionally, regarding authenticity of assessment tasks, the Paper & Pencil format has limitations of content and materials to be administered in a test. Moreover, the P&P format does not take advantage of recent developments in educational measurements such as item response theory and computer-adaptive algorithms for tailored assessment. Therefore, the Department of Applied Linguistics and TESL of UCLA provides WebLAS, a Web-based language assessment system, to overcome the limitations of the P&P tests by specifying the objectives of the project; i.e., to create language test tasks, to administer language assessments, and to increase

flexibility in managing, processing, analyzing, and reporting assessment data to meet the needs of all the language programs at UCLA. Moreover, it aims to guide the design and development of assessment for the ESL, Japanese, and Korean language programs that meet accepted standards of measurement and assessment practices.

It can be concluded that the Paper and Pencil Tests encounter some drawbacks, which make it likely to be replaced by more sophisticated technology tests, the web-based tests in particular.

To follow the current trends in the language test methods, Alderson (2000: 353) states that "the future availability of tests on the Internet will make available a range of media and information sources that can be integrated into the test." Brown (1996) also adds that the Web-based tests become increasingly important and the Web-based approaches are likely to replace the traditional assessment delivery because (a) Web-based tests can interactively yield the results at a high rate without human resource limitations, (b) Web-based tests are convenient as test-takers can get access to the tests regardless of place and time, (c) Web-based tests are interactive because they incorporate interactive features into the tasks, (d) tests can enhance speed of data collection as well as data analyses, and (e) they also yield results on a large scale. On this subject, Röver (2001a) reported his study on Web-based C-tests that C-tests were appropriate to be delivered via the Web because the construction of the tests does not require a great deal of programming expertise and the exact method of scoring, where only one correct answer was accepted, supplied the simplicity of the construction of the C-tests on the Web and this also made the Web-based C-tests objective.

In regard to the advantages of Web-based tests, Röver (2001b) illustrated three advantages of Web-based tests (WBTs): asynchrony, grassroots accessibility, and low cost. To clarify, an asynchronous test does not require the test-taker and the tester to be at the same place at the same time in that WBTs can be taken by any test-taker with an Internet connection and the appropriate software at any convenient time; and the scoring can be done objectively, automatically and immediately. In terms of grassroots accessibility, the WBTs are simple in construction. That is, the creation of the Web-based items is simple to anybody with a computer, an Internet connection, and an introductory HTML book to create a simple Web-based testing program while

the creation of computer-based tests requires advanced and sophisticated technical expertise. Finally, the low cost, WBT delivery costs much less than the delivery of traditional CBTs because WBTs are delivered through public lines and are displayed widely and freely through Web browsers.

Despite the advantages, the WBTs have some limitation compared to those of general computer-based testing; for example, there might be problems with system failure, items security, and exposure. Moreover, some problems might occur due to different levels of test-takers' computer familiarity (Taylor et al., 1998).

Regarding the security of the administration of the Web-based tests, Röver (2001b) explained that the Web-based language testing was a medium-stakes assessment such as placement tests for foreign students, midterm, or final exams in classes, and other assessment situations which did not have board or life-altering consequences to learners' lives and it was likely that test-takers at medium-stakes assessment may have motivation to cheat. Thus, he suggested that in terms of security a placement test had to be administered at a reliable site, preferably in the institution's computer laboratory supervised by a lab monitor to prevent test-takers from misconduct.

In developing a Web-based C-Test, a software program is needed. My Internet search yielded two software projects that employed C-tests as tools of language assessment: one was by Lucy Georges (Georges, n.d) and the other was by Joe Anthony Blum (Blum, 2003). The former was constructed under the name of C-tests 1.2; it provided a construction of C-tests online. However, due to its limitations in constructing passages in the C-test format and also due to no further implication on how the testers employed the tests through the Web-based methods, and since most of the information was written in French, the researcher turned to the latter. It was the http://clozeonline.us site, which implemented Klein-Braley's C-test (1982 cited in Blum, 2003) to establish a training, tutoring, and research institute. Blum's software seems to provide what all language tests should have; that is, ease of construction, administration, and assessment. Thus, with the contribution of the creator of the website, the Web-based C-Test (WBCT) in this research study was constructed, administered, and scored through http://clozeonline.us. In regard to test security, all

teacher tests and scores are in directories protected by a teacher's username and a password.

Apart from studying the feasibility of employing the Web-based C-Test as a placement test for Rajamangala University of Technology Tawan-ok, this study also investigated the test-takers' strategies and the test-takers' opinions towards the Web-based C-Test through a retrospective interview since there has not been sufficient information on how test-takers performed the tests and what test-taking strategies the test-takers implemented in doing the tests. As a result, learning students' test taking strategies in doing the Web-based C-Test may reveal how they process language and the results may help teachers understand how the students employ their language knowledge in doing the test. In addition, the test-takers' opinions towards the Web-based C-Test may provide some empirical evidence towards their preferences for the newly constructed C-test implemented on line, the Web-based C-Test.

In conclusion, placement tests are essential for educational institutions as the tests can place students at their appropriate levels. This can facilitate the teaching and learning process. Rajamangala University of Technology Tawan-ok has no placement tests and the issue of what test should be used by the institution has been raised. This study has explored how effectively the Web-based C-Test could be used in assessing general proficiency. For this purpose, the study has implemented a C-test as a placement test to classify the subjects' English ability levels. In addition, with the web-based delivery, the C-test was constructed by the Internet's software program on http://clozeonline.us. In so doing, the subjects of this study took the Web-based C-Test on the Internet. The subjects were classified into different ability levels based on the Web-based C-Test's scores to indicate that the Web-based C-Test may be used as an English placement test at Rajamangala University of Technology Tawan-ok (RMUTTO). Also, test-takers' strategies and opinions towards the WBCT were explored in this study.

Objectives of the study

The main objectives of the study are:

1. To investigate the effectiveness of the Web-based C-Test in placing the students into different levels according to their English abilities.
2. To explore the test taking strategies of the "high achievers" and "low achievers" in doing the Web-based C-Test.
3. To study the opinions of the "high achievers" and "low achievers" towards the Web-based C-Test.

Research questions

In order to achieve the objectives of the study, this study aims at answering the following questions:

1. Based on the Quick Placement Test, how many levels of English proficiency can the Web-based C-Test (WBCT) differentiate?
2. What are the test-taking strategies that "high achievers" and "low achievers" employ in doing the Web-based C-Test?
3. What are "high achievers' and low achievers' opinions" towards the Web-based C-Test?

Statements of hypotheses

To investigate the effectiveness of the Web-based C-Test, the following hypotheses are formulated:

1. There are significant differences among the mean scores of the students at each level.
2. The students at "high" and "low" levels use different strategies in doing the Web-based C-Test.
3. The students at "high" and "low" levels have different opinions towards the Web-based C-Test.

Scope of the study

This study examined 134 first-year undergraduate students out of 670, which is 20% of the target population. The subjects are from Rajamangala University of Technology Tawan-ok (RMUTTO), Chakrabongse Bhuvanath Campus, in nine majors; namely, 1) Management, 2) Accounting, 3) Marketing, 4) Economics, 5) Information System, 6) Computer Science, 7) Computer Technology, 8) English as an International Communication, and 9) Logistics Technology in the academic year of 2007.

Limitations of the study

1. Due to the fact that the English proficiency of the students of Rajamangala University of Technology Tawan-ok's was limited, only the three low levels of the Cambridge Quick Placement Test were implemented on level classification of the Web-based C-Test.
2. Due to the limitation of the Internet rooms at the institution, the Web-based C-Test was unable to be administered at the same time, but within a week. After each administration, the tests were transferred to another directory for the security of the tests.

Assumptions of the study

Test scores reflect the subjects' actual abilities and the retrospective interview reflects their valid opinions.

Definition of terms

1. **The Web-based C-Test** or **WBCT** refers to the test implemented online through a website, http://clozeonline.us, which is dedicated to the C-test format where the second half of a certain number of second words is deleted.
2. **Undergraduate students** refer to first-year Rajamangala University of Technology Tawan-ok (RMUTTO) students at Chakrabongse Bhuvanath

Campus in nine majors; namely, 1) Management, 2) Accounting, 3) Marketing, 4) Economics, 5) Information System, 6) Computer Science, 7) Computer Technology, 8) English as an International Communication, and 9) Logistics Technology, in the academic year of 2007.

3. **Test-takers' strategies** refer to strategies used in restoring the second half of every second word deleted in the five passages implemented on the Web-based C-Test.
4. **Opinions** refer to students' opinions in doing the Web-based C-Test and their opinions on implementing the Web-based C-Test as an English Placement Test.
5. **The Quick Placement Test** refers to a standardized placement test developed by Cambridge. Test results are reported on the ALTE (Association of Language Testers in Europe) scale. This test is used as a criterion to observe validity in this study.

Significance of the study

1. The students can be placed into different ability levels according to their proficiency so that they can be assigned to appropriate English courses.
2. Knowledge about Web-based C-Test taking strategies is obtained. This can reflect their processing the language and the results can help teachers improve teaching and learning process.
3. The test takers' opinions towards the Web-based C-Test administration are given to reflect students' preferences for the test and they are useful for the test development and further research.
4. Since C-tests are simply tailored to specific needs, other campuses of Rajamangala University of Technology Tawan-ok (RMUTTO); namely, Utentawai, Bang-Pra, and Chantaburi, may gain benefits from the Web-based C-Test. Therefore, a low-cost English Placement Test for Rajamangala University of Technology Tawan-ok (RMUTTO) is achieved.

This research study has combined the uses of the C-tests regarding the C-test format, text selection, and research studies with the Web-based test delivery. This resulted in the Web-based C-Test (WBCT) to be implemented online for data collection and data analysis. The outcomes from the WBCT were analyzed and reported in order to answer the research questions. The use of the WBCT in this study was expectedly implemented as an English placement test for the institution.

Overview

Chapter one presents the background of the study including the objectives, the research questions, the hypotheses, the scope, the limitations, the assumptions, the definition of terms and the significance of the study.

Chapter two portrays a literature review concerning language tests, cloze tests, C-tests, research studies related to C-tests, the construction of C-tests, the construction of WBCT, test delivery methods, computer use in language testing, Quick Placement Test, and test-taking strategies.

Chapter three describes research procedures in terms of the population and sample, research instruments, instrument development, data collection and data analysis.

Chapter four reveals the findings of the study regarding the descriptive statistics of the WBCT and the QPT, the text difficulty indices of the WBCT, the reliability of the WBCT, the correlation between the WBCT and the QPT, and the level classification of the WBCT. Additionally, it presents the content analysis on the test-takers' strategies in doing the WBCT and the opinions of the subjects towards the WBCT.

Chapter five presents the summary, discussion, conclusions, and recommendations for implementations of the WBCT, recommendations for pedagogical implications as well as recommendations for the future research studies.

CHAPTER II

LITERATURE REVIEW

This chapter reviews the literature related to language tests, cloze tests, C-tests, research studies related to C-tests, the construction of C-tests, the construction of WBCT, test delivery methods, computer use in language testing, Quick Placement Test, and test-taking strategies.

Language tests

Language tests can be divided into four types of tests according to their purposes; namely, proficiency tests, achievement tests, diagnostic tests, and placement tests (Hughes, 2003). The four types of tests can be defined as follows:

Proficiency tests

Proficiency tests can be defined as tests that are not related to the content or objectives of any language courses, curriculum, or single skill, but their purpose is to assess test-takers' overall proficiency (Brown, 2004, Hughes, 2003). An example of a standardized proficiency test is the Test of English as a Foreign Language (TOEFL), which measures overall language ability in four skills: listening comprehension, structure (or grammatical accuracy), reading comprehension, and written expression. In order to be able to evaluate overall proficiency, test tasks must represent the language use. Therefore, "creating these tasks and validating them with research is a time consuming and costly process" (Brown, 2004: 45).

Achievement tests

Achievement tests are based on language courses such as classroom lessons or units and their purpose is to assess how successful test-takers achieve the course objectives at the end of a unit or the instruction (Brown, 2004, Hughes, 2003).

Diagnostic tests

Diagnostic tests are used to identify learners' strengths and weaknesses (Hughes, 2003: 15). A diagnostic test can be designed to assess specified aspects such as pronunciation in order to diagnose the test-takers' phonological features (Brown, 2004: 46).

Placement tests

Placement tests are usually used to gather information that is necessary in organizing classes in the teaching program so that students are assigned to appropriate levels (Hughes, 2003). In addition, placement tests help to make decisions to place test-takers into the most appropriate levels (Bachman and Palmer, 1996). Regarding test administration, a placement exam needs to be constructed to support quick administration and the test format should support quick evaluation (Mathews, 1988). Moreover, a placement exam should accurately place all students regardless of the amount of lapsed time since they last studied the language (Mathews, 1988). However, it is not economical to build a big test that can measure overall ability for making decisions that are not of critical importance and backwash is not considered (Hughes, 2003).

The features of the four test types reviewed above provide information that the implementation of each test type depends much on their purposes. Therefore, it is obvious that a test should be thoroughly designed to accomplish the objectives of the tests.

Besides the variations of the language tests on their purposes, there are distinctions between discrete point and integrative testing. The two types of tests can be explained as follows:

Discrete point tests

Discrete point tests refer to tests that are designed to assess one element at a time, to say, item by item as in multiple choice tests. The tests are in the form of a series of items and each item tests a particular language point (Hughes, 2003). The construction of the discrete point tests is based on the assumption that language composes of several components that can be divided into parts and each part can be

successfully evaluated separately (Brown, 2004). Discrete point tests provide no conceptualized information for test-takers to do the tests.

Integrative tests

Integrative tests refer to tests that require test-takers to employ their language ability to combine many language elements to complete the test tasks (Hughes, 2003). The two types of tests, which are claimed to be examples of integrative tests, are cloze tests and dictations (Brown, 2004).

Discrete point and integrative testing lead the concepts of evaluation of small units and overall language proficiency, which still appear in language tests. To see what language tests can be implemented in evaluation, qualities of tests are at all times taken into consideration.

Qualities of language tests

In designing and developing a language test, it is important to consider qualities of the test and the most important quality of a test is its usefulness, which, according to Bachman and Palmer (1996), is a combination of six test qualities as shown in the model (Bachman & Palmer, 1996: 18).

Figure 2.1: Usefulness

Usefulness = Reliability + Construct validity + Authenticity + Interactiveness + Impact + Practicality

In regard to the model of test usefulness, the qualities of a test are controlled by the test usefulness during the development and the implementation of a test; however, it is appropriate for test developers to balance among these qualities, which means that one testing situation may vary from another (Bachman & Palmer, 1996). Also, it is essential to consider the three principles as the basis for the operation of the model of usefulness as follows (Bachman & Palmer, 1996: 18):

Principle 1 It is the overall usefulness of the test that is to be maximized, rather than the individual qualities that affect usefulness.

Principle 2 The individual test qualities cannot be evaluated independently, but must be evaluated in terms of their combined effect on the overall usefulness of the test.

Principle 3 The usefulness and the appropriate balance among the different qualities cannot be prescribed in general, but must be determined for each specific testing situation.

The three principles provide the ideas that a usefulness of a test is considered when it is developed with a specific purpose, a particular group of test-takers and a specific language use domain (Bachman & Palmer, 1996). In regard to test usefulness, it can be perceived in terms of six test qualities: reliability, construct validity, authenticity, interactiveness, impact and practicality. According to Bachman and Palmer (1996), reliability and construct validity are essential for measurement of test qualities while the rest serve pedagogical purposes to promote learning, which are, authenticity in instruction, interactiveness in learning tasks, impact in learning activities, and practicality in teaching approaches.

In the review of language tests, it provides a variation of tests designed upon purposes of measurement. Also, the qualities of the language tests should be taken into account. Among tests, cloze tests have been considered a proficiency test in respect to integrative testing and among quality tests. However, cloze tests have encountered some deficiencies which lead to other test developments.

Cloze tests: the reduced redundancy tests

A cloze test was originated by Taylor (1953) who first developed the cloze procedure as a measure of redundancy. The cloze procedure reduces redundancy by deleting a number of words, usually every seventh word (plus or minus two) as a fixed-ratio deletion, or by deleting grammatical or discourse functions of the words as a rational deletion (Brown, 2004). The test-takers are required to fill in the missing words. In terms of scoring, there are two approaches to the scoring of cloze tests: the exact word method and the appropriate method of scoring. The exact word method accepts the insertion of exact words that were originally deleted while the appropriate

method accepts the insertion of grammatically correct words that make sense in the context (Brown, 2004).

As for the cloze procedure, it is a procedure based largely on the concept of "closure" in Gestalt psychology. The following illustrates some ideas of Gestalt psychology.

Gestalt psychology

Gestalt psychology is, in part, the concept that learners follow the sequence of understanding the whole before getting into individual details. The perception of the whole and the sum of the parts is displayed in Figure 2.2 (Fultz, n.d).

Figure 2.2
The whole is different from the sum of the parts

It is obvious that the whole is different from the sum of the parts. This can be applied to visual perception according to Gestalt principles that "in looking at a tree, we first see it as a single unit, rather than as a collection of individual leaves and branches" (Stansfield, 1980: 715). When applied to cloze tests, the damaged text of a cloze test is perceived by the examinee as a whole gestalt; therefore, the examinee is able to fill in the missing words based on the mental construct or the gestalt, which are established while reading the damaged text (Sigott, 2004). In Gestalt psychology, the perception of the whole is the significant point and is different from the perception of the sum of its parts. Therefore, "humans are able to fill in what is missing by using their background knowledge or their experience," (Heaton, 1975; Oller, 1979 cited in Boonsathorn et al., 2007: 9).

Discrepancies of cloze tests

Cloze tests have encountered criticisms in various aspects. In terms of the qualities of cloze tests, there are five major problems (Raatz & Klein-Braley, 1981: 117):

1. The cloze construction principle (deletion of every n^{th} word) does not ensure random deletion of the elements in the text.
2. Each of the factors (text, deletion rate, and starting point) affects the quality of the test as assessed by the reliability and validity coefficients.
3. The ratio of structure/ content words deleted affects the difficulty of the test.
4. The test constructor cannot tell in advance how his text, his deletion rate and his starting point will affect the characteristics of the test.
5. Tests using the higher deletion rates (8, 9, 10) need extremely long texts if 50 items are to be ensured.

In terms of scoring, cloze tests have high discrepancies between the exact method and acceptable method of scoring. The problems of scoring are displayed as follows (Raatz & Klein-Braley, 1981: 118):

1. Exact scoring, while objective, is inappropriate for L2 learners unless they have received training in completing cloze tests.
2. Acceptable scoring is not objective.
3. The scoring method used affects the performance of the test.
4. The two methods of scoring are not necessarily equivalent for advanced L2 students.

The discrepancies of cloze tests in the qualities and the scoring methods have led to attempts of improving cloze tests. A modified cloze test that employed the concept of closure in Gestalt psychology and the theory of reduced redundancy in cloze test procedure is the C-test. It is, therefore, correct to perceive C-tests as a modification of cloze tests.

C-tests: a modification of cloze tests

According to Raatz and Klein-Braley (1981), a C-test is a language test of proficiency based on the concept of reduced redundancy. A reduced redundancy theory states that a redundant message contains more information than is necessary to understand the message. Thus, when some parts of the message are damaged or omitted, the left message can still convey the understanding of the message. In C-tests, the second half of every other word is deleted leaving the first halves as clues for restoration. This means the undamaged parts make the restoration possible to complete the message. However, a C-test does not assess discrete points. It is, therefore, an integrative written test (Raatz and Klein-Braley, 1981) as test-takers employ their integrative abilities to do a C-test. It is not the redundancy of the text that is measured, but the test-takers' ability to employ their language ability to restore the missing parts of the text (Raatz and Klein-Braley, 1981).

In brief, there are five or six short authentic texts of which each passage completes in a sense unit by itself. The first sentence in each text is not deleted. The "Rule of 2" is employed to delete the second half of every second word starting at the second sentence of the text. The deletion does not affect numbers and proper names. As for words with an odd number of characters, the deletion is on the larger part. One-letter words such as 'a' and 'I' are not counted. The deletion continues until a certain number of deletions are reached (either 20 or 25); thereafter, the texts are allowed to complete the paragraph.

To illustrate the different formats between a cloze test and a C-test, the following excerpts show the fixed-ratio deletion of a cloze test and a C-test procedure.

Example 2.1

The fixed-ratio deletion of a cloze test and a C-test procedure

Cloze procedure, fixed-ratio deletion (every seventh word)	**C-test procedure**
The recognition that one's feelings of (1) _____and unhappiness can coexist much like (2) _____and hate in a close relationship (3) _____offer valuable clues how to (4) _____a happier life. It suggests, for (5) _____, that changing or avoiding things that (6) _____ you miserable may well make you (7) _____ miserable but probably no happier. (Brown, 2004: 202)	The recognition th_ _ one's feel_ _ _ _ of happ_ _ _ _ _ and unhap_ _ _ _ _ _ can coe_ _ _ _ much li_ _ love a_ _ hate i_ a cl_ _ _ relati_ _ _ _ _ _ may of_ _ _ valuable cl_ _ _ on h_ _ to le_ _ a hap_ _ _ _ life. I_ suggests, f_ _ example, th_ _ changing o_ avoiding thi_ _ _ that ma_ _ you mise_ _ _ _ _ may we_ _ make y_ _ less mise_ _ _ _ _ but prob_ _ _ _ no hap_ _ _ _. (Brown, 2004: 204).
Full text: *The recognition that one's feelings of happiness and unhappiness can coexist much like love and hate in a close relationship may offer valuable clues how to lead a happier life. It suggests, for example, that changing or avoiding things that make you miserable may well make you less miserable but probably no happier.*	

The strategic "Rule of 2"

The term that best defines a C-test is the "Rule of 2," which is the deletion technique provided to organize the test. The "Rule of 2" refers to deleting every second multi-character word, excluding proper names, starting with the second word of the second sentence. Multi-character words imply the exception of the "Rule of 2" that the word has to be more than one character in length. Hastings (2002a) gives an explanation to this technique that the reason to start deleting from the second sentence is similar to cloze tests in that the first sentence provides test-takers with information about the topic and style of the passages and the other "two" of the rule characterize the C-testing in that the deletion on every second word promotes the representativeness of the sample and the half word deletion makes the test relatively

simple to restore. He provides this demonstration of why half of the words rather than whole words are deleted (Hastings, 2002a:14):

> However, if every second word is completely deleted, the resulting test will be too difficult, as the following example suggests:
>
> **I _____ you _____ agree _____ exact _____ is _____ .**
>
> With exact scoring, few native speakers would be able to get a respectable score on such a test: there are too many possible solutions in addition to the actual original words.
>
> However, if only the second half of every second word is deleted, the test is fairly easy for a native speaker of English, especially if the number of deleted letters is indicated:
>
> **I th_ _ _ you wi_ _ agree th_ _ exact sco_ _ _ _ is desir_ _ _ _ .**

In addition, Hastings (2002a) explains that there is enough redundancy in the first half for a proficient person to reconstruct the missing part correctly and only one word is correct so the exact method of scoring is used.

The reduced redundancy theory

C-tests employ the same principles to cloze tests. According to Taylor (1953), the creator of cloze tests, cloze tests are based on Gestalt psychology and information theory. The two principles of the cloze procedure have been known as related to reduced redundancy testing (Sigott, 2004).

The concept of reduced redundancy has played major roles in language testing and it has entered applied linguistics through readability research (Sigott, 2004). Since Taylor (1953) proposed the cloze procedure as a measure of readability, there were studies to investigate whether cloze tests were based on the principle of reduced redundancy or not. Spolsky (1971, 1973 cited in Sigott, 2004: 18) attained the answer that "natural language contains redundancy in order to safeguard a message against noise which may interfere as the message is transmitted from a sender to a

receiver." In so doing, the message transmitted is still comprehensible despite omission of some parts because every message carries many elements to make the message understandable. Sigott (2004: 18) explained that "knowing a language is not equivalent to knowing the discrete elements, but it also involves knowing about the redundancies inherent in the system." Thus, he concluded that the ability to understand a message, of which the redundancy was reduced, assessed the overall language proficiency of the speaker.

As for C-tests, which are proficiency tests underlying the reduced redundancy theory (Raatz and Klein-Braley, 1981), the notion of reduced redundancy derives from the basic assumption that adult native speakers who are educated and familiar with a language can complete the missing parts by taking advantage of the redundancy of the language. In so doing, they employ their knowledge of the rules, patterns and idioms as well as culture to demonstrate their competence. This indicates that the tests of reduced redundancy evaluate overall language proficiency. Thus, "the less proficient a speaker, the more redundancy he or she will require for successful communication" (Sigott, 2004: 18). This can be perceived by the fact that a less proficient learner of a language does not have sufficient competence to make use of redundancies to restore the message. Raatz and Klein-Braley (1981) also point out that in restoring the words in C-tests, the learner's ability of making use of the redundancy of the language is measured, not the redundancy of the text itself. This indicates that C-tests do not measure discrete units, but general proficiency.

It can be perceived that the theory of the reduced redundancy behind the C-test help learners to restore the missing parts. The better the proficiency of the learners, the more restoration they can produce. As such, the perfect scores will inevitably reflect the learners' proficiency pertinent to native speakers. Essentially, redundancy is a feature of natural language. It is, therefore, common that the second half of the words deleted in C-test format appears in the test to induce general proficiency of the test-takers to restore the words. Common sense then dictates that parts of a message can be distorted or missing yet still communicates meaning. Competent users of the language will simply understand the message.

Like cloze tests, C-tests are based on the reduced redundancy which constituted from the Gestalt psychology and information theory (Taylor, 1953). It can be seen that the concepts of perceiving things as a whole is applicable in the C-tests. Thus, the examinee has to employ background knowledge and experience in restoring the damaged parts. It is, therefore, not the damaged parts that are measured, but the examinees' proficiency to employ their knowledge by perceiving things as a whole. The fact that the damaged parts contain both content words and function words determine that examinees have to employ general proficiency in restoring the missing parts. This can evaluate the examinee's integrative language proficiency. For example, the restoration of the word 'i_' requires the examinee's understanding of the whole context whether the restoration should be 'is', 'if', 'it' or 'in'. If it is for discrete unit assessment, the examinee may use the guessing method in restoring the missing part, instead. Thus, learners are to integrate their knowledge as a whole to restore the missing parts in order to reconstruct the original passages. Apparently, the theory of reduced redundancy and the Gestalt principle applied to C-tests provide information that is adequate to ensure users of the C-tests that the tests generally assess learners' integrative proficiency, not just discrete units.

In conclusion, both cloze tests and C-tests are successful because (a) they measure the extent to which learners have managed to mature to see the Gestalt of mutilated texts and have, by so doing, learned to restore the missing characters; and (b) they take advantage of the fact that most culturally rich languages have a great number of redundant words in each sentence and paragraph. Thus, both cloze and C-tests reduce that redundancy without losing the text's meaning. In shorter words, the Gestalt theory and the Reduced Redundancy theory underlie the cloze tests and C-tests. Both tests are built on these two principles.

What C-tests measure

Initially, in 1981, C-tests were introduced to teachers in Germany as classroom achievement tests. However, when Klein-Braley (1994 cited in Raatz & Klein-Braley, 2002) investigated whether C-tests and cloze tests measure general proficiency, she confirmed that the two tests measure the same construct and that C-

tests were superior to cloze tests. Moreover, Dörnyei and Katona (1992, 1993) indicated that the C-test had a correlation with the other language-proficiency tests in vocabulary, grammar scores and speaking skills. Also, Rashid (2004) indicated that the C-tests measure general language proficiency. Consequently, Klein-Braley and Grotjahn (Klein-Braley & Grotjahn, 1995 cited in Raatz & Klein-Braley, 2002), recommended using C-tests as a proficiency test because C-tests did not assess discrete units and the assessment was not related to school curriculum. It is, therefore, appropriate to implement C-tests as proficiency tests rather than achievement tests. Moreover, Eckes and Grotjahn (2006) investigated what C-tests actually measured. The findings yielded that the C-test was a highly reliable, unidimensional instrument, which measured the same four skills as in the TestDaF (Test of German as a Foreign Language); reading, listening, writing and speaking. In particular, more constructs were divided by the language proficiency and the different level of proficiency affected the C-test performance.

Whether C-tests assess high- or low-level processing was inclusive. However, according to many research studies, C-tests investigated high-level processing. Connelly (1997) investigated C-test as a measure of general language proficiency of high level students studying English as a foreign language. The results indicated that the C-test was highly reliable and has concurrent validity in assessing the English proficiency among EFL contexts. Sigott (2002) conducted a study to investigate high-level processes in C-test taking. The results indicated that about half of the items involved high-level processing. Also, Hastings (2002b) investigated error analysis of an English C-test to find evidence for integrated processing and the results revealed that in doing the C-test, the test-taker employed their knowledge of semantic, syntactic, morphological, lexical, and orthographic processing in restoring the missing parts. This also indicated the high-level processing.

It can be concluded from the studies reviewed above that the high-level processing in doing the C-tests seems to be pertinent to the characteristics of the test-takers who have high language proficiency. It is, therefore, interesting to investigate students in Thai context, whose language proficiency is limited, to find out what C-

tests measure in this context and how the test-takers employ their knowledge in doing the test. Particularly, whether the students in the tertiary level employ their integrative competences to do the tests should be taken into account.

Criticisms of C-tests

Like cloze tests, C-tests encounter some problems. Grotjahn et al. (2002) point out some major problems with C-tests: (a) C-tests tend to be too simple for very advanced learners so the discrimination of advanced learners is limited. (b) For a norm-referenced test, C-tests are found reliable even when mean of difficulty is below 50 per cent. And (c) C-tests are criticized as related to reading and writing learned at school and there are no oral tests. The justification of this criticism is related to the reduced redundancy theory, stating that all language users have the ability to make use of the natural redundancy which is an asset rather than a valid point of criticism.

There are also problems with the deletion technique, which occurs during the implementation of C-tests in some languages. For example, the deletion technique worked well for English while there were problems with German because of the polymorphemic character of many German words (Grotjahn et al., 2002). In addition, a similar problem appears with Italian and French words. Italian words with apostrophes must be treated as two words while French words beginning with a letter plus an apostrophe and words joined by hyphens should be treated as one word (Grotjahn, 1987).

Also Carroll (1987) adds some critical remarks about C-tests, saying: (a) When using the C-test in testing native language skills, it tends to center on reading and writing skills; (b) the positive correlations obtained between the C-test scores and teachers' judgment of language competence is suspected as teachers may have too narrow concepts of language competence; (c) with regard to using C-tests to assess second language skills, C-test has been pointed out to measure 'crossword puzzle solution skills' instead of language competence; (d) the lack of agreement on how C-tests should be administered may also result in the 'leak' about the deletion of the

second half of words to potential candidates and it is likely to distort the results; (e) the word ‘and’ occurs several times and is always omitted and this would make the test too easy; (f) the lead in the sentence at the beginning of each text and a full ending sentence often provide clues for examinees as it is likely that a deleted word in the text may reappear at the end of the text; and (g) C-test procedure is also criticized to be arbitrary with limited uses.

C-tests also encounter face validity problems. Legenhausen (1989 cited in Sigott, 2004) studies the face validity of a battery of C-tests in secondary schools’ Germany Language Competition using C-tests together with other tests like multiple-choice test of grammatical structures and communicative competence, a writing task, a listening comprehension test, and a test of oral skills. The data were collected from the teachers involved in administering the tests. The results revealed that most teachers considered the C-tests too difficult for their students although the test results showed that the C-tests were easy for the candidates (Grotjahn 1989, Raatz & Klein-Braley, 1989 cited in Sigott, 2004). Legenhausen assumed that the teachers’ judgments of excessive difficulty were twofold: first, the teachers’ random samples of the test’s disadvantages were collected from students who had less structural or lexical competence; and second, the teachers were not familiar with the test format. Legenhausen added that some teachers comment on the lack of validity of the C-test and that the test lacks authenticity since solving the blanks involves blind guessing.

The criticisms over C-tests varied in the implementation of other languages, ranging from its deletion technique to test administration and face validity. More empirical evidence may illustrate the strengths and weaknesses of C-tests so that solutions can be found. It is anticipated that some of the problems raised above might also occur in Thai context. For example: C-tests may be criticized as being a test of reading and writing skills, not general proficiency; C-tests may be considered a test of ‘crossword puzzle solving skills’ instead of language competence; the ‘leak’ about the deletion of the second half of words may occur and is likely to distort the results; the lead in the sentence at the beginning of each text and a full ending sentence often provide clues for examinees as it is likely that a deleted word in the text may reappear at the end of the text; and C-tests may encounter face validity problems.

Research studies related to C-tests

Since Raatz and Klein-Braley (1981) initiated the use of C-tests as an alternative in assessing language ability, there are research studies attempting to validate its measurement of general language proficiency in other target languages. Dörnyei and Katona (1992, 1993) validated the C-test among Hungarian EFL learners. The subjects were 102 Hungarian university students and 53 Hungarian secondary students. The C-test, the Department Proficiency Test, TOEIC, Oral Interview, and Cloze Test, were administered to the subjects and the results of the five tests were compared. The findings indicated that the C-test was a reliable testing instrument as it correlated well with the other language-proficiency tests. It was highly correlated with vocabulary and grammar scores, as well as speaking skills. The C-test measured general language proficiency more accurately than the cloze test. They concluded that the C-test was an appropriate testing method to measure language proficiency.

Likewise, Ikeguchi (1998) conducted a study to discriminate the levels of language proficiency between ESL learners in Japan using the two types of C-tests: C-test 1 was constructed from several passages and C-test 2 from one long passage or the Narration. The subjects were two groups of freshmen university students in Japan: 60 undergraduate students enrolled in the general Freshman English course and 30 students in Freshman English for returnees. The STEP exam was used as the external criterion. The results indicated that ESL levels of students in Japan could be differentiated by the two types of C-tests implemented in the study and that C-test 1 was more reliable and valid than C-test 2.

Moreover, Rashid (2004) did a study to validate the C-test as a measure of the English language proficiency among Malay ESL learners. The subjects were 30 secondary Malay male students representing two academic levels: Form One and Form Four. The 14 subjects in the Form One group studied English for at least six years while the 16 subjects in the Form Four group studied English for at least ten years. The two sets of C-test were English C-test 1 and English C-test 2. Both were composed of four sub-tests and were constructed from the same texts. Each subtest

had twenty-five items. The second form of the C-test differed from the first form in that the first word deleted was made at the third word leaving the others at the second word deletion. The second form of the C-tests was administered three weeks after the first administration. The subjects were given 45 minutes to complete the tests. Two different cloze tests that were used as the placement test for their English proficiency were used as the criterion measure. It was found that the English C-tests in this study yielded high reliability (.80 and above). The mean difficulty indices for both forms of the C-test ranged from .60 to .69 in both sample groups. The validity of both English C-tests and the correlation coefficients were high, ranging from .94 and .96, suggesting that the two tests were measuring the same construct; that is, general language proficiency.

Eckes and Grotjahn (2006) investigated what C-tests actually measured. The subjects were 843 participants from four independent samples who took a German C-test and the TestDaF (Test of German as a Foreign Language). The findings yielded that the C-test was a highly reliable, unidimensional instrument, which measured the same four skills as in the TestDaF (Test of German as a Foreign Language); reading, listening, writing and speaking. In particular, they showed that language proficiency was infinitely more complex and that its variances influenced the C-test results.

There were other research studies attempting to modify the C-test in terms of different words deletions. Boonsathorn (1987) initiated the Modified C-test or MC-test, where the first half of the words is deleted to compare with the C-test, where the deletion occurs at the second half of the words. Two forms of the C-test and two forms of the MC-test were administered to 389 native English speakers and 104 ESL learners. The ESL learners also took the Michigan Test, which was used as a criterion measure, and 28 of them were chosen for oral interviews. The results indicated that the MC-test discriminated better and had a greater reliability and validity than the C-test. The MC-test required more reading strategies than the C-test. The interviews revealed that the high proficiency group and the low proficiency group used similar strategies in processing the C-test/ MC-test passages despite their different reader types. The researcher claimed that the MC-test was superior to the C-test.

Likewise, Boonsathorn (2000) initiated the S-Tests or the Semantic/ Syntactic Tests which delete only content words. The tests were created on the basis of some weaknesses pointing out that the "grammatically unmarked items" or the "function word items" are far too easy (Cleary,1988; Dörnyei & Kotona,1992 cited in Boonsathorn, 2000). The S-Tests are comprised of S1-Test and S2-Test in that the S1-Test deletes the first half of every second content words while the S2-Test deletes the second half of every second content word. Boonsathorn states that because most content words contain both semantic and syntactic information in themselves, the term S-Tests (Semantic/ Syntactic Test) are used. The sample was 97 graduate students at the National Institute of Development Administration (NIDA). Two parallel forms of S-Tests, the S1-Test and the S2-Test were developed. The findings showed that both the S1-Test and S2-Test had high reliability. The S1-Test had higher discrimination power than the S2-Test. From the questionnaire, the respondents agreed that both tests were interesting and challenging language tests and that the S1-Test was considered more challenging than the S2-Test. Additionally, the respondents agreed that the S1-Test measured mainly vocabulary, background knowledge and general English proficiency while the S2-Test measured mainly vocabulary, reading and general English proficiency. In conclusion, both tests had proven to be reliable and valid tests. The S1-Test appeared to be superior to the S2-Test in reliability, concurrent validity, face validity, and difficulty suitability.

Prapphal (1994) did a study on the C-test and the X-Test (the reversed C-test) performed by first year science-oriented university students of Chulalongkorn University: 38 engineering students and 18 science students. Passages from The Bangkok Post were selected and processed into both types of tests. Scoring was on the basis of the "exact word" method. The findings indicated that both the C-test and the X-Test were highly reliable. When the C-test was presented before the X-Test, there was no significant difference; however, a significant difference was found when the X-Test was presented before the C-test.

Moreover, Rungruangtham (2005) investigated four types of cloze tests, namely, C-test, NC-test (or the new C-test of every third word deletion), MC-test (the reversed C-test), and NMC-test (the new reversed C-test of every third word

deletion), by comparing the mean scores of the C-test with those of the NC-test, and the mean scores of the MC-test with those of the NMC-test in measuring the English language proficiency among 110 first-year undergraduate science students in the Faculty of Science, Mahidol University. The results indicated that there were no statistically significant differences between the C-test and the NC-test in both high and low language ability groups. There was a statistically significant difference between the MC-test and the NMC-test in the high language ability group. It was concluded that the NC-test may be used as a substitute for the C-test in measuring English proficiency of EFL tertiary students. The MC-test and the NMC-test were acceptable to be used for measuring English proficiency of high language ability students while the NMC-test was less suitable for measuring low-language-ability students than the original MC-test. The findings suggested that the teachers intending to implement the four cloze tests should take deletion techniques and the selection of the text contents into serious consideration.

Additionally, Lertudommit (2005) conducted a study to investigate whether the S1-Test (the Semantic/ Syntactic Test of the first half of every second content word deletion) and the S2-Test (the Semantic/ Syntactic Test of the second half of every second content word deletion) could be used as measures of general English proficiency of grade eight students at an international school. The subjects were 80 students enrolling at a large international school in northeastern Bangkok. The S1-Test and the S2-Test were constructed from the same reading passages with the readability levels ranging from grade 5.8 to 8.5. The Quick Placement Test was initially administered to the subjects to assess their general English proficiency. The findings revealed that both the S1-Test and the S-2 Test had high reliability and moderate concurrent validity. The difficulty levels of both the S1-Test and the S2-Test were appropriate for grade eight students and both tests had discrimination power. The subjects had good attitudes towards both tests and also agreed that the tests measure general English ability.

Boonsathorn et al. (2007) developed NIDA Test of English Competence (N-TEC) for the National Institute of Development Administration's graduate students and the general public. A new format of Cloze Test, the Semantic/ Syntactic Test or

S-Test, was used as a section in the battery of tests. The N-TEC was comprised of three sections: Error Identification, the S-Test, and Reading Comprehension. The subjects were 209 voluntary participants consisting of graduate students at NIDA and the general public. Two forms of the N-TEC were used with 105 participants completing Form A and 104 completing Form B. The results indicated that both forms of the N-TEC had high reliability and high criterion-related validity. The data from the questionnaire assessing the face validity supported that the N-TEC had high face validity. The findings revealed that N-TEC Form A was suitable for the target group while N-TEC Form B might be a little too difficult. Discrimination power was 66% of the items in Form A and 67% of the items in Form B.

Other investigations of C-tests focus on whether high- or low-language abilities are measured. Connelly (1997) investigated the C-test as a measure of general language proficiency of high-level students studying English as a foreign language. The subjects were non-native postgraduate students of engineering, technology, and management attending pre-master's programs from six different countries, namely, Thailand, Vietnam, Taiwan, Indonesia, Japan, and Cambodia, at the Asian Institute of Technology (AIT) in Bangkok, Thailand. The English C-test with 100-item deletion was implemented. The results indicated that the C-test was highly reliable and had concurrent validity in assessing the English proficiency of students in EFL contexts.

Likewise, Sigott (2002) conducted a study to investigate high-level processes in C-test taking. The results indicated that a great number of subjects employed high-level processing in doing half of the items while few chosen items necessarily required high-level processing. He concluded that a lot many of the subjects employed high-level processing in dealing with both easy and difficult items from all four word classes and C-tests were not tests of lower-order skills.

Further, Hastings (2002b) conducted a study to investigate error analysis of an English C-test to find evidence for integrated processing. Error analysis was performed on a large corpus of C-test responses from 200 international university students in several intensive English as a Second Language programs in the United

States. These students had taken a C-test as a part of the placement and promotion procedures in their program. The C-test for the study consisted of 12 short paragraphs on a variety of topics, chosen randomly from sources like newspapers, magazines, and popular fiction. These patterns indicated that the C-test assessed the ability to integrate contextual information with a range of language competencies, including those involved in semantic, syntactic, morphological, lexical, and orthographic processing.

Babaii and Moghaddam (2006) conducted a study to investigate the relationships between test-task difficulty and macro-level procession in the C-test. The subjects were 119 English-major students; 36 of these students participated in the retrospective think-aloud phase. Four texts with low and high degrees of syntactic complexity and abstraction were selected. Two sets of the C-test with the same content were composed: one with and one without providing clues regarding the number of missing letters. The analyses revealed that employing texts with more syntactic complexity and abstraction without clues of the number of missing letters could result in more difficult test tasks that seemed to encourage more frequent use of macro-level processing.

Regarding research studies attempting to validate the C-test with standardized tests, Daller and Phelan (2006) investigated whether the C-test could partially replace the Test of English for International Communication (TOEIC) and whether the C-test was an appropriate format for the measurement of students' progress in short intensive language courses. The TOEIC was used as a selection and placement instrument. The subjects in the study were first-year undergraduate students from France in the UK who took a TOEIC exam before and after an 11-week intensive English course with 240 hours of teaching. An entry and an exit C-test were administered. The study showed that both tests were sufficiently reliable and that both tests measured a significant increase during the course. Based on this study, the C-test was accepted and implemented as the economically best choice for measuring the progress of large numbers of students.

It can be seen that the original C-test has led to a number of research studies in regard to its strengths in evaluating proficiency. Many studies had attempts at modifying the C-test to suit specific purposes, yet virtually all agreed that the C-test concept is a valid approach to measuring language proficiency.

However, there are investigations that question the effectiveness of the C-tests. Jafarpur (1995) investigated the possibility of using a C-test with native and non-native speakers of English. The instruments of this study were the C-test in 20 versions of different deletion frequencies and deletion starting points, and one standard cloze. The subjects were 202 native English speakers who studied in various fields at the University of New Mexico and 325 non-native speakers of English who studied in the Department of Foreign Languages and Linguistics at Shiraz University in Iran. The findings revealed that the C-test was simple in the construction and scoring; however, it suffered the same discrepancies found on the classical cloze procedure, particularly in validity.

Moreover, Jafarpur (1999) performed another study to explore the effectiveness of the "Rule of 2" in C-testing using five versions of a C-test and a standard cloze test with 340 Iranians majoring in English. The C-tests were constructed with three different deletion starts and two different ratios. The results showed that the varying deletion rates and deletion starts yielded more or less similar results. Towards Jafarpur's findings, Hastings (2002a) reported in his defense of C-testing under the attack by Jafarpur saying that Jafarpur not only misinterpreted the claims made for C-tests but also failed to put those claims to a fair test. Hastings pointed out that varying two parameters violated the rules for C-test construction and Jafarpur's tests could not be considered C-tests because Jafarpur's experimental tests were all based on a single passage, which constituted another violation of the rules for C-test construction.

Rouhani (2008) conducted a study to investigate the validity and discrimination power of the C-test to assess overall language proficiency with 144 Iranian EFL university students. The results indicated that despite the C-test's high reliability and high criterion-related validity to a Michigan Test of English Language Proficiency (MTELP), the C-test texts did not have consistency to differentiate the

examinees' proficiency levels and that the C-test could not consistently classify the subjects in their appropriate proficiency levels. Particularly, the C-test had difficulty discriminating between participants at lower and upper intermediate levels.

Even the opposition to the use of C-tests has led to meaningful investigations. Jafarpur's and Rouhani's studies made variations to the C-test studies that proved that their variations could not consistently classify the EFL learners in Iran into appropriate levels. It is, therefore, interesting to investigate whether the Web-based C-Test will prove the same with EFL learners in Thailand or whether it will do what it does in other countries when used properly.

Construction of C-tests

In order to construct a C-test, it is essential to follow the principles of the C-test, criteria for the construction of C-tests and principles of text selection.

Principles of C-tests

In 1981, Raatz and Klein-Braley set up the principles for the constructions of the C-test as follows (Raatz and Klein-Braley, 2002: 74):

1. The new test should be much shorter, but at the same time it should have at least 100 items.
2. The deletion rate and the starting point for deletions should be fixed.
3. The words affected by the deletions should be a genuinely representative sample of the elements of the text.
4. Examinees with special knowledge should not be favoured by specific texts, therefore the new test ought to consist of a number of different texts.
5. Only exact scoring should be possible so as to ensure objectivity.
6. Native speakers ought to be able to make virtually perfect scores on the test: 90% or higher. If native speakers cannot make scores higher than 90%, then the text should not be used for non-native speakers.
7. The new test should be reliable, valid and easy to develop.

Criteria for the construction of C-tests

Raatz and Klein-Braley inaugurated the classical C-tests' criterion in construction of C-tests in order to ensure that C-tests are objective, reliable and valid to measure general language proficiency. The following steps are necessary (Raatz and Klein-Braley, 2002: 84):

1. Define the target population and the test format.
2. Choose suitable texts, more than necessary, using regression equation for predicting difficulty.
3. Bring the texts into C-test format and combine them into one or more C-tests.
4. Examine the tests in a group of educated adult native speakers.
5. Analyze each text: Is the mean difficulty 90% or more? Are there gaps in the text with more than one solution?
6. Decide either: text is satisfactory; OR text is satisfactory, but some words have to be changed; AND/OR in some damaged words letters have to be added; OR text is not satisfactory, i.e. too difficult.
7. Combine the good texts – possibly after correction – into one or more C-tests in order of ascending estimated difficulty.
8. Examine the C-test in a larger representative sample of the target population.
9. Perform item analysis, estimate test reliability and validity (if a criterion is available).
10. Perform CLA-analysis in order to investigate the dimensionality and the type of scale of the test.
11. Decide: C-test is satisfactory, i.e. it is acceptably reliable and valid, unidimensional, and interval scaled; OR C-test is approximately satisfactory, but some texts or the order of texts have to be changed, or a text has to be excluded; OR C-test is not satisfactory (in this case one has to start all over again).
12. Improve the test, construct the final form.
13. Perform additional studies for reliability and validity.

14. Administer the final form to a large representative sample of the target population.
15. Calculate the test norms on the basis of the distribution of the raw scores.

Principles of text selection

Raatz and Klein-Braley (2002: 85) point out that in selecting appropriate texts, the following principles should be taken into account. The texts should be:

1. complete in themselves;
2. appropriate in difficulty and content for target group;
3. with no specialized vocabulary and content;
4. no literary texts, no verbal humor.

Furthermore, Raatz and Klein-Braley (2002) add other essential points in text selection: (a) possible sources of texts are non-fictional books, newspapers, magazines, brochures, and information leaflets; (b) some authentic texts might be too difficult for EFL learners, so quasi-authentic texts that are specially written by native speakers of the language for EFL learners are advisable; (c) texts that are equivalent to text books are recommended; and (d) some minor alternations in the authentic texts are also possible.

It is apparent that to find appropriate texts to meet the principles of text selection is sometimes time-consuming in that the selected texts should be able to ascertain test fairness. According to Klein-Braley and Raatz (1984), text selection should not benefit any examinee to employ specialized knowledge in doing the test. Grotjahn and Stemmer (1985) also conclude that test-takers employ more strategies to restore content words rather than function words in difficult texts, which indicates that more difficult texts require the solution of a C-test in semantic processing. As perceived, text selection is an essential process that should be taken into careful consideration.

Construction of the WBCT

The C-test can be developed differently depending on the purposes it serves. In reviewing C-tests and the construction of C-tests, the researcher was able to constitute criteria for the construction of the WBCT and criteria for text selection in this study, which was particularly implemented as an English placement test for RMUTTO's students in Thai context. The criteria for both the construction of the WBCT and text selection were adapted from those of the classical C-tests to suit the implementation of this study.

Criteria for the construction of the WBCT

The purpose of the construction of the WBCT in this study is to implement the constructed test as an English placement test for the RMUTTO. Therefore, the criteria for the construction of the WBCT are as follows:

1. Define the target population and the test format.
2. Choose suitable texts using Flesch-Kincaid Grade Level's Readability in Microsoft Word for text difficulty. The appropriate Grade Level's Readability ranges from Grade 5 to Grade 12.
3. Ask three raters to rate on the texts in terms of these characteristics: general in contents, authentic either by native writers or in sources, accurate in language use, appropriate and interesting to teenagers (first-year students).
4. Use the software program to construct the texts into the C-test format and combine the five subtests into the Web-based C-Test (WBCT). The software program transforms each text into the C-test format in four versions: (a) what the students will see, (b) the paper version with blanks around the underscores, (c) the full text with deletion markers, and (d) the colorized hidden words. The colorized hidden words help the researcher to check on the appropriateness of the deleted parts.
5. If the deletion is not satisfactory, decide whether: (a) some words have to be changed, (b) some damaged words or letters have to be added, or (c) some texts have to be discarded because of their difficulty.

6. Try out the WBCT with three native speakers. Native speakers ought to be able to make 90% or higher scores on the test.
7. Try out the WBCT in Pilot Study with a sample of the target population.
8. Perform item analysis to obtain facility index (IF) and discrimination index (ID). If any of the texts obtains either IF or ID lower than 2.00, the texts should be replaced.
9. Estimate test reliability and concurrent validity with a criterion test.
10. Improve the WBCT and construct the final form.
11. Perform additional studies for reliability and validity.
12. Administer the final form to a representative sample of the target population.
13. Analyze the data.

Criteria for text selection of the WBCT

The WBCT in this study employed the classical C-test criteria, including some recent research findings, and set the WBCT criteria to be as follows:

1. The WBCT consists of five short passages. Each passage consists of 60-100 words and it should complete as a sense unit by itself.
2. The passages come from a variety of sources to avoid content bias.
3. The passages should be general, authentic, and interesting to the target groups.
4. There is no specialized vocabulary and content.
5. There are no literary texts and no verbal humor.
6. Text difficulty is appropriate to the target group; the readability of the passages is between Grade 5 and Grade 12.
7. Each passage has 20 deletions and the passage is allowed to run on to its conclusion after the 20 deletions.
8. The WBCT has 100 items. Only exact scoring is employed to ensure objectivity. One point is assigned to each correct answer.
9. If a word has an odd number of letters, the larger "half" is deleted.
10. Words that consist of only one letter, numbers, and proper names are left undamaged in the deletion process.

11. Some minor alternations in the authentic texts are also permitted.
12. Native speakers ought to be able to make 90% or higher scores on the test.

The adapted criteria for the construction of the WBCT would be appropriate to Thai learners who learn English as a foreign language (EFL) and have limited language proficiency. Therefore, the selection of the texts has to be performed thoroughly in order to meet all the prescribed criteria.

The software program

In order to organize the WBCT, a software program is needed to transform the short reading passages into the WBCT. The Web-based C-Test was constructed with the software program at http://clozeonline.us. The engine of the project corresponds to Klein-Braley's C-test (1982 cited in Blum, 2003), which has been considered an efficient tool for assessing language proficiency. The software program not only performs the deletions as prescribed (starting at word two in sentence two and then the second half of every second word until the required number of 20 blanks is produced) but also provides a method for dealing with proper names, which need to be left intact, and paragraph titles.

The software program, at ClozeOnline.us, is a program developed by Joe Anthony Blum. Blum (2003: 3-4) lists these three aspects of ClozeOnline™:

1. ClozeOnline™ is an assessment instrument that uses the C-test (Klein-Braley, 1982), now administered online, to evaluate (a) foreign language teachers in their own country, (b) foreign language teachers in a foreign country, (c) teacher competency in using a foreign language to teach a subject, and (d) to test the degree to which students have become native-like.
2. ClozeOnline™ is achievement-measuring software that produces classic cloze tests (Taylor, 1953), L-tests (Kokkota, 1988), and several versions of Klein-Braley's C-test. It can be installed on a school's own server or used

online through private directories. The program also creates a printable version of the tests.

3. ClozeOnline™ is the engine used for training language skills at all levels starting with grade 2, specifically the elements of (a) active reading, (b) reading comprehension, (c) paragraph-style text creation, (d) use of proper grammar, and (e) meaning-based text restoration. On a peripheral level, taking tests on a weekly basis has proven to enhance a student's test-taking skills.

With the software program, the passages are organized into the WBCT. There are two different processes in organizing the WBCT, which are the process of making the tests for instructors and the process of doing the tests for students or test-takers (For more details, see Appendix A: 121).

The process of making the WBCT for instructors

The process of making the tests was adapted from the http://clozeonline.us site, the program used for this study. There are seven steps in constructing the WBCT as follows:

1. Go to "http://clozeonline.us for teachers" and type in the user name and password, which are supplied by the webmaster, then press 'LOG ON.'
2. At the "Create New Test" screen, name the test with two strings, e.g., a course and page code, then click on "Create New Test."
3. At the "Make Test" screen, the name of the test appears on the screen. Then, copy the selected passages from a file and paste them into the box, with this string >>> between each passage so that the passages will appear to the students randomly one at a time.
4. Change the defaults to a variety of combinations or leave the settings at second word deletion with the number of word deletions set to 20 and click 'MAKE TEST.'
5. In details, at the http://clozeonline.us for teachers, the instructions can be read as follows:

Example 2.2

Instructions for making the WBCT

- New Test. Type or paste your text into the box, then click **MAKE TEST** below.
  Type >>> between separate subtests (paragraphs).

- If no word lists are chosen, every second word (of more than 1 character) will be truncated.
  Select the number of words to
  mutilate:
  Select NUL (number of undeleted characters): 50%+ 50%- 1 0
- Academic word list(s): **1 2 3 4 5 6 7 8 9 10**
- Conventional list of words: **1 2 3 4**
- Research list(s): **1 2 3**
- Use just my own list of words: **1**
- Encrypt the hidden parts of words in the -missing.txt file:

Place an asterisk * at the start of a word to prevent it from being hidden.
For details on creating a classic cloze, L-test and C-tests, click **HERE**.

6. After clicking on 'MAKE TEST,' the test appears in 4 different versions; that is, a) what the students will see, b) the paper version with blanks around the underscores, c) the full text with deletion markers, and d) the colorized hidden words.
7. In the meantime, the newly constructed C-tests are automatically listed on the http://clozeonline.us for students' screen. The test is now ready for students to log on and do the tests.

The process of doing the WBCT for test-takers

The process of doing the test for test-takers was adapted from the software program to be implemented in this study. The test-takers are required to follow the following steps:

1. The student logs on to "http://clozeonline.us for students" and enters his/her name at [Student Name] followed by an email address in the format of "name@name.com."
2. This is followed by entering the [Teacher's Name], in my case, "NADD" for the sample test and "CPCTEST" for the actual test, the WBCT," and clicking [LOGON].
3. When the student clicks on the name of the test, in my case, "SAMPLE-official" for the tutoring period or "WBCT-official" for the actual test, the test appears one paragraph at a time.
4. Before doing the test, the student needs to follow the instructions on doing the WBCT. The researcher stresses that the student presses the keyboard's [Insert] key and activate (click) the curser inside the text box before using the arrow keys to maneuver the curser to the beginning of the first underscore. The instructions can be read as follows:

Example 2.3

Instructions for doing the WBCT

CPCTEST: WBCT-official

AFTER pressing [Insert], replace the missing characters.
USE [ARROW KEYS] ONLY! If you use [Del] or [Backspace], press [Esc] and start over.
Switch to Full Screen [F-11] to see the text below the [Submit] button.

5. For the sample test, the researcher acts as a coach to familiarize the student with the test format and the process in doing the WBCT.
6. The student clicks on 'no' to the statement under the test that says: *I am a native speaker of English: yes no.*
7. The student is given 9-10 minutes to restore each text before clicking on the SUBMIT button. The results after submission of the test appear at the bottom as displayed:

Example 2.4

The results on screen after submission

> Your score is currently 8 out of 20. Replace more underscores and **SUBMIT** again,
> or **skip ahead to the next test in this series,**
> **or end the tests now. (*You must click to get credit.*)**
> **Thank you for aiding our research.**

8. For this study, students get only one try, therefore, the student clicks on *'Skip ahead to the next test in this series'* at the bottom of the screen. Another test appears. This process repeats until the student finishes the five subtests provided.
9. As the student submits each test, the results are displayed on the screen as well as sent to the researcher's e-mail.
10. When working with this online version of the C-test, it is important to stress the pressing of the [Insert] key after every time the SUBMIT button is pressed. Should they forget (which is evidenced by seeing text being pushed out of the way), they could use the [Esc] key to restore the window to its proper integrity.
11. The time allotment for the WBCT is 45-50 minutes for restoring the five subtests.

Test delivery methods

Tests are delivered in methods varying from the traditional paper-based to the web-based delivery methods. The methods vary to serve the purposes of the tests. Researchers are to assign appropriate test delivery methods to facilitate data collection. There are three main test delivery methods that are frequently used among teachers and researchers: the traditional paper and pencil or paper-based tests, the technological computer-based tests, and the hypertext Internet- or the web-based tests. The practical uses of the three methods are given below.

Paper-based Tests (PBT)

The most classical and frequent test delivery used either in the classroom or for research studies seems to be the paper-based tests or PBT. Not only the school or university tests are in the form of paper-based test, but most of the standardized tests obtained at university testing or other services are performed on a basis of paper and pencil tests; for example, the GRE, GMAT, TOEFL, SAT, IELTS, to name a few.

The paper-based test methods are utilized in many types of tests such as cloze tests and tests originated from the cloze test. Since Taylor (1953) initiated the cloze test as an instrument for assessing the relative readability of written materials for school children in the United States, Oller (1979) and Cohen's (1980) utilized cloze tests as a measure of overall ESL proficiency and the tests were originally administered on paper-based tests. A number of studies were designed on PBT to explore cloze tests' reliability and validity. It can be concluded that cloze tests utilize the PBT as a test delivery and so do modified cloze-tests and other tests.

Computer-based Tests (CBT)

Due to new trends of test delivery methods, the computer-based test or CBT has become increasingly popular and there is the feasibility that the CBT will replace the PBT. This is because computers have higher capacities in administering and scoring. Egbert and Handson-Smith (1999) suggest that in computer-assisted language (CAL) classrooms, technology can play two overlapping roles in the process of assessment in that both process and outcomes are assessed. In addition, the availability of the CBT involves various factors including teacher performances and objectives of the tests. As a result, computer-based tests can be performed according to its usefulness and its potentials. The simple basic problems that occur in computer-based tests are the accessibility of the computers, devices, software and electricity.

To give an example, the TOEFL computer-based test measures English language proficiency in three disciplines: reading, listening, and writing. The Listening and Structure sections of the TOEFL computer-based test (CBT) are computer-adaptive, meaning; they start with a question of average difficulty. If the first question is correctly answered, the computer will give a question of higher

difficulty. So, based on the answers, the computer can evaluate the level of English and provide questions according to the test-taker's level.

Internet-based Tests (iBT) or Web-based Tests (WBT)

The services on the Internet have provided users with a variety of activities including language testing. The Internet-based Tests are used for classroom instruction, research studies as well as language proficiency tests. The Internet-based Test (iBT) and the Web-based Test (WBT) are the advanced uses of the CBT; however, the iBT or WBT can get closer to participants around the world regardless of the capacities of the computer, place and time.

Basically, a web-based test (WBT) is written in the Internet language, HTML, and is delivered through the internet in order to provide test-taker's accessibility to the test, which is located at a website on the tester's server. This cannot be performed with the traditional CBTs that need platforms to deliver the tests. Apparently, WBTs are cheaper in production and maintenance. However, there are problems in the implementation of WBTs in the identification of test-takers and the prevention of item downloading and other exposure as well as whether the test-takers do the tests privately in their own home (Röver, 2001a).

To give an example, there is a new testing format called DIALANG (Alderson: 2005) which provides tests for anyone learning a language in reading, writing, listening, grammar and vocabulary. This project is developing diagnostic language tests in 14 European languages delivered over the Internet; they are, Danish, Dutch, English, Finnish, French, German, Greek, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish and Swedish. Furthermore, it provides a wide range of feedback and advice. The test system is based on the Common European Framework and incorporates self-assessment and detailed feedback on performance and advice on how to improve one's language proficiency. In other words, it finds out the test-takers' levels of strengths and weaknesses as basic information to make decisions on how to develop their mastery of a language. Tests of reading, listening, writing, grammar and vocabulary are freely available at www.dialang.org. In addition, "there are a number of web-based tests (WBT) available on the internet, for example,

QuestionMark, Cambridge, IELTS, TOEFL, TOEIC, Computerized placement tests (LOPT, TOPE)" (Prapphal, 1997: 124).

In regard to research, researchers can effectively administer and evaluate tests through the Internet. Moreover, the iBT increases paces of data collection as well as data analysis. In the C-test performance, Röver (2001a) implemented the tests through the web and suggested that the creation of the tests did not need a great deal of programming expertise.

It can be perceived that the WBT is for anybody with access to the Internet. It is operated on the basis that a web browser can complete the test at any time from anywhere in the world. It is likely that the Web-based delivery method will become more popular and more effective than the computer-based test method. However, there are limitations that require development to ensure more uses of the Web-based delivery and the security of the tests. Besides, some limitations are relevant to the context of Thai classrooms, which are the simple basic problems that occur with computer-based tests, problems such as lack of devices, software programs, electricity, and access to the Internet. Additional problems could be caused by operator error and system failure. However, following the international trend to continually improve conditions and processes will be the key solution to most of these problems.

Computer use in language testing

The computer has played a very important role in testing, especially in test delivery. Despite some shortcomings, there are a number of advantages to using computers in language testing. Cohen (1994) points out that computer-based testing has implemented modern technology in that it uses the computer for assessment. The new technology has brought great changes to test delivery and computer-based tests (CBT) are replacing the paper-based tests. This supports the convenience in creating the CBT under supported programs. Additionally, the CBT seems to be more exciting and more advanced to the test-takers because the feedback is dynamic. Also, computer software assists teachers to generate a number of computer-based tests.

The advantages of using computers in language testing can be divided into two categories: testing considerations and human considerations (Brown, 1992). Among the testing considerations of using computers, the advantages are on computer accuracy and immediate feedback. Among human considerations, advantages include students' allowance to work at their own pace. Moreover, the efficiency of Computer Assisted Language Tests (CALTs) is greater than that of the traditional paper-and-pencil tests in that students taking CALTs are reported to be less frustrated than those taking paper-and-pencil tests because the questions are presented one at a time on the screen rather than the whole test in a booklet (Brown, 1992). In addition, many students like computers and enjoy the testing process (Stevenson & Gross, 1991).

There are also disadvantages in using a computer in language testing. Brown (1992) points out that the disadvantages of using computers in language testing can be divided into two categories: physical considerations and performance considerations. Among the physical considerations, he lists problems with lack of equipment and reliable sources of electricity as well as limited screen capacity. The screen size limitations could be a problem for a reading test based on long passages. In addition, many computers have limitations in their graphics capabilities. Among the performance considerations, the disadvantages include the presentation of a test on a computer, which may lead to different results if the same tests were administered in a paper-and-pencil format (Henning, 1991 cited in Brown, 1992). Besides, "the students' different degrees of ability in using computers or typewriter keyboards may lead to inconsistency in their performances on computer-assisted or computer-adaptive tests" (Hicks, 1989; Henning, 1991; Kirsch, Jamieson, Taylor, & Eignor, 1997 cited in Brown, 1992). Another disadvantage includes the potential effects of computer anxiety on test performance (Henning, 1991 cited in Brown, 1992).

In regard to research, Kantahoe, Hansombatcharoen, and Leardvitoonpanya (2003) proposed a computer-based S-Test to cope with the inconvenience of the test scoring and the lack of test attraction to test-takers that occurred with paper-and-pencil based test. However, Maulan (2004) examined the effects of computers as a test delivery medium for English language test performance in a writing test. The results revealed that the subjects performed significantly better in a paper-and-pencil

writing test than in a computer-assisted test. There was no significant relationship between the subjects' computer-familiarity and performance on computer-assisted tests.

Likewise, Kiratibodee (2006) studied the relationships between three selected variables and Computer-Based Test (CBT) performance in testing reading comprehension. The three variables were computer attitudes, computer anxiety, and computer familiarity. The findings indicated that computer attitudes and computer anxiety were significantly correlated with CBT scores of students at all levels of language ability. Additionally, computer familiarity was a significant predictor of CBT scores of students with average language ability, while computer attitudes were a significant predictor of CBT scores of students with high language ability.

As for the situation of Thai classrooms in using computer-based testing, there are meager research studies that encourage the implementation of the computer for web-based tests. This study, aiming to implement a C-test via a Web-based delivery, may encourage teachers and researchers in carrying out more studies in using computers for language assessment.

Quick Placement Test (QPT)

The QPT is designed to help teachers make decisions about the class to place students in, or whether a learner can join a particular course, such as an examination class, and it can be used by learners of all levels and all ages (Quick Placement Test Paper and Pen Test User Manual, 2001: 2). The results are linked to the ALTE levels, the Common European Framework (CEF), and Cambridge ESOL Exams. The QPT is not designed to be used as a progress test, as it is not based on a particular course or syllabus.

ALTE Levels

The QPT reports test results as a band on the ALTE (Association of Language Testers in Europe) 6-level scale. The following table illustrates the look-up table for computer-based and paper and pen scores.

Table 2.1

Look-up table for computer-based and paper and pen scores.

(Quick Placement Test: Paper and Pen Test User Manual, 2001: 8)

ALTE Level	**Paper and Pen test score**		**Computer-based test score**
	Part 1 Score out of 40	Part 1&2 Score out of 60	
0 Beginner	**0 - 15**	0 -17	0-39
1 Elementary	**16 - 23**	18 – 29	40 – 49
2 Lower Intermediate	**24 - 30**	30 – 39	50 – 59
3 Upper Intermediate	**31 – 40**	40 – 47	60 – 69
4 Advanced	**If a student scores 36 or more, it is recommended they complete Part 2 of the test.**	48 - 54	70 – 79
5 Very Advanced		55 - 60	80 - 100

Table 2.1 shows how to interpret the results of each version of the test in terms of the ALTE Levels from 0 to 5, and so to compare computer-based scores against paper and pen scores. It is recommended that only Part 1 of the paper and pen test is used if the students are mostly between ALTE Levels 0 and 3. Additionally, Part 1 and Part 2 should be used when the level of the students is unsure or when students score more than 36 on Part 1 of the test, or when the teacher believe that the students are Level 3 or above. Therefore, there are two columns for interpreting scores on the paper and pen tests. Column 2 shows the band scores for students taking Part 1 with the maximum score of 40, covering ALTE Levels 0 to 3 while column 3 shows the band scores for students taking both parts of a paper and pen test with the maximum of 60, covering the range of ALTE Levels from 0 to 5. If students take the second part of a paper and pen test, they will get 20 more difficult items than those in Part 1.

The key features of the computer-based version are that the test is adaptive, so different students see different questions. In addition, the test assesses Listening, Reading, Vocabulary and Grammar. All the questions in the test are in multiple-choice format. The test is marked by the computer and the result is generated immediately. The result can be compared with the result of the paper and pen version (Quick Placement Test: Paper and Pen Test User Manual, 2001: 2).

The key features of the paper and pen version are that the test assesses Reading, Vocabulary and Grammar. It takes 30 minutes to administer and all the questions in the test are in multiple-choice format. The test consists of two parts. Part 1 is taken by all students. Part 2 is for higher ability students only. The result can be compared with the computer-based result (Quick Placement Test: Paper and Pen Test User Manual, 2001: 2).

In terms of standard error of measurement or SEM, for the computer-based test the margin of error is ± 4 points on the 100-point scale; for the 60-item paper and pen test it is ± 4 points; and for 40-item paper and pen test it is ± 3 points of their true score (Quick Placement Test: Paper and Pen Test User Manual, 2001: 9).

In addition, ALTE (Association of Language Testers in Europe), which is an association of providers of European foreign language examinations, has included some of the major international providers in the field of language testing; that is, the Council of Europe Level and the Cambridge Examinations. The following table presents the chart of equivalent levels of the ALTE, the Council of Europe and the Cambridge Examinations.

Table 2.2

Chart of Equivalent Levels

(Quick Placement Test: Paper and Pen Test User Manual, 2001: 9)

ALTE Level	ALTE Level Description	Council of Europe Level	Cambridge Examinations
0	Beginner (Breakthrough)	A1	
1	Elementary (Waystage)	A2	KET
2	Lower Intermediate (Threshold))	B1	PET BEC Preliminary CELS Preliminary
3	Upper Intermediate (Independent user)	B2	FCE BEC Intermediate CELS Intermediate
4	Advanced (Competent user)	C1	CAE BEC Advanced CELS Higher
5	Very Advanced (Good user)	C2	CPE

According to the Quick Placement Test: Paper and Pen Test User Manual (2001: 10), the salient features of the ALTE levels are as follows:

Breakthrough Level: a basic ability to communicate and exchange information.

Level 1 (Waystage): capacity to deal with simple straightforward information, and to get by in familiar contexts.

Level 2 (Threshold): can mange a wider range of situations; express opinion and attitude in a limited way.

Level 3 (Independent User): confidence, useful functional ability, greater awareness of conventions of use.

Level 4 (Competent User): fluency, adaptability, sensibility, capacity to deal with the non-routine.

Level 5 (Good User): a degree of linguistic virtuosity, capacity to work quickly; deal with cognitively demanding material, use language to effect.

To sum up, the Cambridge Quick Placement Test, QPT, has potentials in placing examinees into appropriate levels in accordance with the ALTE scale, the Council of Europe Level, and Cambridge Examinations. In this study, the computer-based version was employed in the pilot study 1 and the paper and pen version was implemented in the pilot study 2 and the main study to examine the concurrent validity to the newly constructed Web-based C-Test (WBCT).

Test-taking strategies

Test-taking strategies refer to learner strategies or mental operations that learners consciously select for use in accomplishing language tasks. Their successful use depends first on whether they are appropriate for the given task (Cohen, 1994: 119). It is, therefore, significant that the results of the tests as well as the process that learners used to generate the answers be examined because the strategies used tend to happen during the testing session (Cohen, 1998). Moreover, learners may use the strategies to provide them with short-cuts to the answers. For example, a study of forty college ESL respondents using retrospective verbal report to investigate their test-taking strategies (Larson, 1981 cited in Cohen, 1984) indicate that the students answer a ten-item multiple-choice test based on a 400-word reading passage using several strategies answering items through shortcuts. Moreover, learners may improvise strategies that are appropriate for their own consequences; for example, in a study of test-taking strategies in Israel, one Hebrew second-language respondent wrote a translation of a text before he could answer the questions (Cohen & Aphek, 1979 cited in Cohen, 1984).

In order to get into the realm of test-taking strategies, means for investigating test-taking strategies, including reading strategies and cloze testing strategies, are taken into account.

Means for investigating test-taking strategies

Means for investigating test-taking strategies have been implemented in a number of research methods in order to access and describe the strategies in which the methods have often produced indication or clues about learners' strategies rather than instances of actual strategies. The methods being used are presented below (Cohen, 1994: 122):

1. Questionnaires and interview (e.g., Cohen & Robbins, 1976; Oxford & Crookall, 1989)
2. Observation schedules (e.g., FOCUS; Fanselow, 1977)
3. Performance analysis (e.g., identification of communication strategies for speech protocols, Faerch & Kasper, 1983; and identification of reading strategies from miscue analysis, Goodman, 1988).
4. Language assessment instruments (e.g., cloze testing for which completion of each blank presumes the use of a given strategy, Homburg & Spaan, 1981; Bachman, 1985; Jonz, 1990).

Reading strategies

Reading strategies provide helpful guidelines on how to approach the reading passage. They help the reader organize the passage into main and supporting ideas for a better understanding of the test. Scanning and skimming for relevant cues such as number and capital letters make reading less threatening and more readable. In addition, there is a relationship between general knowledge, skills in applying reading strategies, and reading comprehension (Kozminsky & Kozminsky, 2001:187). Reading becomes more effective when students have background knowledge (McNamara et al., 1996: 248). Students become nervous when they have to read about cultural topics with which they are unfamiliar (Saito et al., 1999: 214).

Hosenfeld (1977 cited in Cohen, 1994) initially conducted research on the reading strategies of nonnative readers and developed a system to record each strategy that the learners used as they read. The findings indicate that while reading competent foreign-language readers keep the content of the passage in mind, read in broad phrases, skip unimportant and unknown words and use the remaining words as clues, and, at a final alternative, look up a word in the back of the book or in a side glossary. These strategies are listed below (Hosenfeld, 1977 cited in Cohen, 1994: 123):

1. Analyzing
2. Translating
3. Attending to grammar
4. Using sentence-level context
5. Using passage-level context
6. Stopping at an unknown word
7. Skipping an unimportant/ inconsequential word or phrase
8. Looking up a word in the back of the book or in a side glossary
9. Viewing the importance of words equally or differently
10. Going back

In Boonsathorn's study (1987), an investigation of 28 students' reading strategies was examined through interviews. It was found that eight students employed reading word by word while 10 students read the whole sentence, and the remaining 10 read the whole text first. When comparing the high-proficiency-group (HPG) and the low-proficiency-group (LHG), the three strategies between the two groups showed not much difference. The finding revealed that students in both proficiency levels employed sentences or the whole passage context to find answers. The low proficiency students did not solely focus on each individual word or the orthographic input. Half of the students implemented translation in doing the test; particularly, the low proficiency students used translation strategies more than the high proficiency students.

Likewise, Nevo (1989 cited in Cohen, 1994) did a study on test-taking strategies of reading comprehension with a checklist of 15 strategies, each appearing

with a brief description and a label to promote rapid processing of the checklist. The first three entries are listed below (Nevo, 1989 cited in Cohen, 1994: 126):

1. Background knowledge: general knowledge outside the text called up by the reading in order to cope with written materials.
2. Guessing: blind guessing not based on any particular rationale.
3. Returning to the passage: returning to the text to look for the correct answer after reading the questions and multiple-choice alternatives.

Reading strategies are very useful in helping test-takers doing reading comprehension tests, in other words dealing with reading passages. Cloze tests and C-tests utilize the redundancies we find in reading passages. Thus, it is probable that the reading strategies may be applied as strategies in cloze tests and C-tests.

Cloze testing strategies

A cloze is the traditional format that has a passage with every n^{th} word fully deleted. In completing the blanks in cloze tests, some researchers state that completing the blanks successfully indicates reading comprehension at the macro level (Chihara, Oller, Weaver, & Chavez-Oller, 1977; Chavez-Oller, Chihara, Weaver, & Oller, 1985 cited in Cohen, 1994) while other researchers point out that filling in the blanks successfully indicates micro, sentence/word-level reading skills rather than macro, discourse-level reading skills' strategies (Alderson, 1983; Klein-Braley, 1981; Bachman, 1985; Lado, 1986 cited in Cohen, 1998). Whether test-takers use macro or micro level processing in doing cloze tests is not conclusive. The following illustrates some ideas with the micro-level and the macro-level strategies (Cohen, 1994: 142):

1. Micro-level strategies
   a) word-level/part of sentence
   b) sentence-level
2. Discourse or macro-level categories
   a) inter-sentential level
   b) whole-text level
   c) extra-textual level

d) metacognitive level

e) others

Similarly, Bachman (1985) developed a framework dividing cloze items into four categories according to the textual information that students implemented while answering each cloze item. They are:

1. Within Clause,
2. Across Clause, Within Sentence,
3. Across Sentences, Within Text,
4. Extratextual.

In Bachman's 1985 study, cloze tests are implemented with different deletions to examine language skills with 910 native and non-native college students in Illinois. The results revealed that the subjects frequently used the 'Extratextual' strategy for the fixed-ratio cloze test, while 'Across Sentence, Within Text' strategy was implemented by students taking the rational cloze test.

Furthermore, in regard to Bachman's framework, Sasaki (2000) conducted a study to examine the effects of cultural schemata on Japanese EFL students responding to unfamiliar and familiar fixed-ratio cloze passages with 60 Japanese EFL students that had similar English reading proficiency level. The students were asked to report their test-taking strategies as categorized based on Bachman's cloze test-taking strategy framework. In addition to Bachman's framework, Sasaki added more test-taking strategies in her study based on the discourse analysis in the previous cloze test-taking framework as follows:

1. Within Clause: The examinee uses information provided only by the clause in which an item appears.
2. Across Clause, Within Sentence: The examinee uses information provided by a broader context than the clause in which an item appears, but a narrower context than the orthographic sentence.

3. Across Sentences, Within Paragraph: The examinee uses information provided by a broader context of orthographic paragraphs containing an item.

4. Across Paragraphs, Within Text: The examinee uses information provided by the context of the entire text.
5. Extratextual: The examinee uses information that is not provided by the text itself, but which is assumed to be included in the examinees' world knowledge.
6. Guessing: The examinee guesses the answer.
7. Missing: The examinee does not/cannot say anything about his or her test-taking processes, or does not answer the item.

Yamashita (2003) examined test-taking strategies based on Sasaki's framework using a rational deletion cloze test and think-aloud protocols with 12 Japanese students. Six skilled readers and six less skilled readers were examined to find whether the cloze test could measure "higher order processing ability." The findings exhibited that both skilled and less skilled readers employed the 'Across Paragraphs, Within Text' strategy more frequently than other strategies. Also, the skilled readers employed the "Across Sentences, Within Paragraph" and the "Across Paragraphs, Within Text" more than the less skilled readers. Yamashita concluded that the rational deletion cloze test could measure English language proficiency of EFL students.

It can be concluded that the cloze testing strategies deal with learners' proficiency in word, clause and sentential levels as well as reading strategies such as guessing. It is apparent that the cloze testing strategies can be applied as C-test strategies since C-tests are modified cloze tests.

The WBCT-taking strategies

Since the C-test is one form of cloze tests, the test-takers' underlying reading strategies and cloze testing strategies stated above were examined in this study. That is, the WBCT-taking strategies, derived from a combination of Bachman's test-taking

framework (1985), Cohen's micro strategies (1994), Nevo's test-taking strategies (1989 cited in Cohen, 1994: 126), and Hosenfeld's reading strategies (1977 cited in Cohen, 1994: 123), were extended for a retrospective interview. The final WBCT-taking strategies in this study are as follows:

1. Reading the clause (Within Clause)
2. Reading the sentence (Across Clause, Within Sentence)
3. Scanning the whole text (Across sentences, Within Text)
4. Reading each word or part of sentence (word-level/part of sentence)
5. Using extra knowledge (Extratextual)
6. Using background knowledge
7. Guessing
8. Returning to the passage to find answers
9. Analyzing clauses, sentences, or the text
10. Translating words or sentences in the context
11. Attending to grammar
12. Others

It can be perceived that test-taking strategies are essential in learning how test-takers respond to the tests. It is obvious that test-takers perform differently because of their differences in abilities and strategies in doing the tests. In other words, the process of doing the test is different among test-takers. The WBCT-taking strategies may help the researcher to explicitly learn the test-takers' strategies in doing the WBCT.

In summary, all the detailed literature of C-tests illustrates the overall picture of how a C-test can be developed theoretically and practically. The theories behind the development of C-tests provide some guidelines that C-tests could be based on a basis of the redundancy theory, through strategic deletions, and the Gestalt psychology of perceiving things as a whole. Together, they convey the concept of test-takers' employing different types of background knowledge to fill in the missing parts. The other part of the literature review also explains the construction of the C-tests and the process of text selection, along with the software program that led to the construction of the WBCT. In addition, this review covers research studies related to C-tests, test

delivery methods and computer use in language testing by stressing the significance of administering C-tests through the Internet and using computers as a medium in terms of convenience and security. The strengths and weaknesses of those studies were taken into consideration while the construction of the WBCT in this study was being developed. Last, test-taking strategies were reviewed so as to provide some guidelines for further exploration of how test-takers manipulate their overall knowledge to cope with the test. Together, a preview of C-tests, the theories behind C-tests, criteria for the construction of C-tests, text selection, the software program, research studies related to C-tests, test delivery, placement tests and test-taking strategies highlight the advantages of creating an affordable test for placing students into appropriate levels at the onset of their training.

Framework of the study

The following figure conceptualizes the framework of the study. It consists of 4 stages as follows:

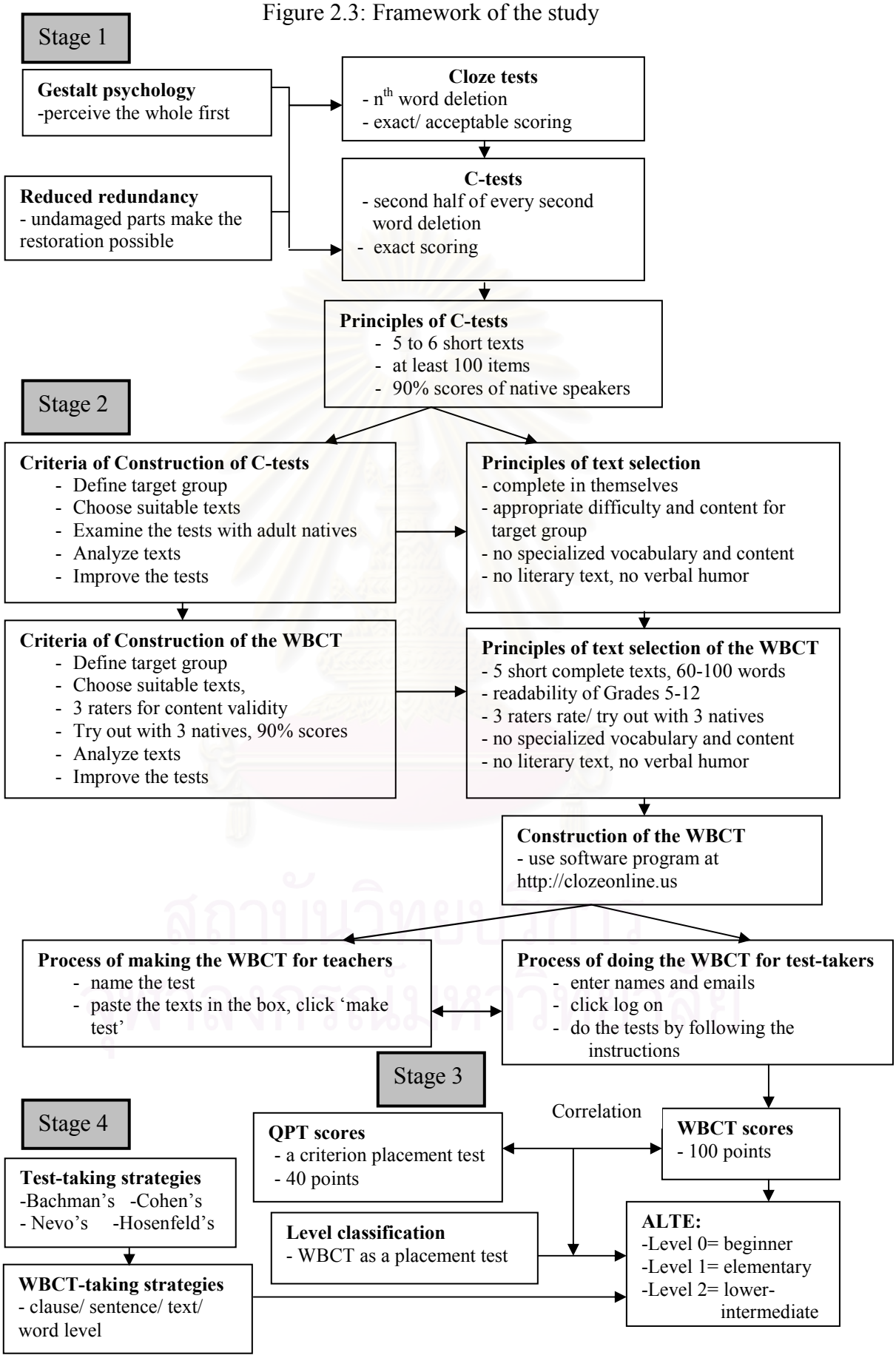
Stage 1: The collaboration of the theories underlying cloze tests and C-tests, the principles of C-tests, the criteria for the construction of C-tests, and the principles of text selection was performed.

Stage 2: The criteria for the construction of C-tests and the principles of text selection led to the criteria of the construction of the WBCT and the principles of text selection for the WBCT performed with that software program including the process of making the WBCT for teachers and the process of doing the WBCT for students.

Stage 3: The scores of the WBCT were correlated to the scores of the criterion test, the QPT, and were calculated for the level classification using levels in ALTE.

Stage 4: The WBCT-taking strategies was employed to investigate the test-takers' strategies of the high and low achievers.

Figure 2.3: Framework of the study



CHAPTER III

RESEARCH PROCEDURES

This chapter describes the research procedures for the selection of the subjects, research instruments, instrument development, data collection, and data analysis.

Population and sample

The subjects of the main study were 134 first-year undergraduate students out of 670, in the academic year of 2007, which was 20% of the target population. They were from nine majors; namely, 1) Management, 2) Accounting, 3) Marketing, 4) Economics, 5) Information System, 6) Computer Science, 7) Computer Technology, 8) English as an International Communication, and 9) Logistics Technology. Thus, the stratified random sampling method was used to obtain the sample for this study. That is, the proportion of students in each major was calculated to obtain the number of the sampling representing the population.

Table 3.1 shows a breakdown of students' programs and number of students in the main study. There are 134 subjects or 20% of the target population in the study.

Table 3.1

Students' programs and number of students in the Main Study

Students' programs	Number of students	
	Total	Main Study
1. Management	99	20
2. Accounting	43	9
3. Marketing	87	17
4. Economics	81	16
5. Information System	62	12
6. Computer Science	47	9
7. Computer Technology	79	16
8. English as an International Communication	93	19
9. Logistics Technology	79	16
Total	670	134

According to Table 3.1, the 134 subjects are from nine majors; that is, 20 from Management, 9 from Accounting, 17 from Marketing, 16 from Economics, 12 from Information System, 9 from Computer Science, 16 from Computer Technology, 19 from English as an International Communication, and 16 from Logistics Technology.

Research instruments

The instruments of this study consist of two tests and a retrospective interview.

1. The available instrument is the Quick Placement Test: Computer-based version and Paper and Pen test Version 1.

   a) The computer-based version was implemented in Pilot study 1. The test is adaptive and it assesses Listening, Reading, Vocabulary and

Grammar. All the questions are in multiple-choice format. The test is marked by the computer and the result is generated immediately.

b) When implemented the computer-based version in Self Access Center at the institution, some computers could not display images and some could not provide listening. The subjects had to move to other computers. This caused confusion and was time consuming.

c) When conducting the Pilot Study 2, the researcher decided to implement the QPT paper and pen version 1 to reduce the problems occurred. According to the manual of the QPT (2001), the two tests can be used interchangeably.

d) As for the paper and pen version 1, it consists of two parts: Part 1 consists of 40 questions and Part 2 consists of 20 questions. If a student scores 36 or more, it is recommended they complete Part 2 of the test. The full score of Part 1 and Part 2 together was 60.

e) The paper and pen version Part 1 and Part 2 were implemented in the Pilot study 2 and the time allotment was 50 minutes. However, in Part 1, no one obtained the score more than 36 out of 40. Therefore, Part 2 was not taken into account.

f) Due to the limited proficiency of the subjects, the paper and pen Quick Placement Test, Part 1, was performed in the main study. And the time allotment was reduced to 30 minutes.

2. The developed instruments are the Web-based C-Test and a retrospective interview. The following shows the procedures of the development of the instruments.

a) As for the Web-based C-Test, 5 out of 10 short reading passages were selected from three experts' ratings in terms of generality, authenticity, accuracy, appropriateness, and interest to first-year students (See the questionnaire for three raters in Appendix B: 129).

b) As for a retrospective interview, there were five questions provided to elicit the test-takers' strategies underlying the WBCT-taking strategies

and another five questions to elicit their opinions towards the WBCT. It was a semi-structured interview. Also, there were questions to encourage the test-takers to include additional perspectives. (See the retrospective interview form available for the researcher in Appendix C: 139).

Instrument development

The developed test (WBCT) was validated in the two pilot studies following the subsequent steps:

1. Select 10 short reading passages with general topics from a variety of sources based on word counts and readability grade levels.
2. Invite three experts to rate the characteristics of the passages employed for the construction of the Web-based C-Test.
3. Select five passages out of ten passages justified by the three raters' International Congruence Agreement for the construction of the Web-based C-Test.
4. Construct the Web-based C-Test using the software program on http://clozeonline.us on the Internet. Each passage contains 20 half-word deletions. There were 100-item deletions to restore.
5. Try out the WBCT on line with three native speakers in order to investigate the basic assumptions that native speakers ought to be able to make virtually perfect scores on the test. If not, the passages should be adjusted or eliminated.

There were two pilot studies to validate the WBCT.

Pilot Study 1

The objective of Pilot Study 1 was to validate the five passages selected from Flesch-Kincaid Grade Level's Readability in Microsoft Word. The subjects of this study were 32 first-year undergraduate students out of 514 (6.23%) using the stratified random sampling method. The subjects were from Rajamangala University

of Technology Tawan-ok (RMUTTO) in Chakrabongse Bhuvanath Campus in the 2006 academic year. They were from eight majors; namely, Management, Accounting, Marketing, Economics, Information System, Computer Science, English as an International Communication, and Logistics Technology. The following table displays the students' program and number of students in the Pilot Study 1.

Table 3.2

Students' programs and number of students in Pilot Study 1

Students' programs	Number of students	
	Total	Pilot Study 1
1. Management	77	5
2. Accounting	35	2
3. Marketing	67	4
4. Economics	58	4
5. Information System	43	3
6. Computer Science	72	4
7. English as an International Communication	76	5
8. Logistics Technology	86	5
Total	514	32

As for the text selection, the five selected passages were carefully selected in accordance to the criteria of the text selection for C-testing. Also, the Flesch-Kincaid Grade Level's Readability in Microsoft Word was taken into account. The following table shows the topics, word counts, readability and sources of the passages in Pilot Study 1.

Table 3.3

Topics, word counts, readability and sources of five passages in Pilot Study 1

Topics	Word counts	Read-ability	Sources
1. TIME IS PRECIOUS	100	5.6	Gordon, G. M. "Teen Give Thanks" *Reader's Digest*. Retrieved from http://www.rd.com/content/openContent.do?contentId=15304
2. DEEP BREATH-ING	95	6.4	Health, S. "The Right Breath", *Reader's Digest*. http://www.rd.com/content/openContent.do?contentId=16354
3. LAUGHTER	88	7.0	"Enhance Your Sense of Humor." *Reader's Digest*. http://www.rd.com/content/openContent.do?contentId=16125
4. TREES	97	9.4	Zimmerman, S. "Plant a Tree in Your Community." *Reader's Digest.* Retrieved from http://www.rd.com/content/openContent.do?contentId=29483
5. MOVIE REVIEWS	71	11.3	Reading Movie Reviews." *Bangkok Post*. http://bangkokpost.net/education/movies.htm

According to the table, the readability of the five passages ranging from 5.6 – 11.3 were selected from the Reader's Digest and the Bangkok Post; that is, Passages 1-4 were from the Reader's Digest and Passage 5 was from the Bangkok Post. The passages in full texts and in C-test format are displayed in Appendix D (See Appendix D: 143).

As for the process of data collection and data analysis, the following displays the steps in collecting and analyzing the data in Pilot Study 1:

1. Permission was granted from the Rajamangala University of Technology, Tawan-Ok, Chakrabongse Bhuvanath Campus to implement the tests.

2. The two tests: the Web-based C-Test (WBCT) and the computer-based Quick Placement Test (QPT) were administered to 32 subjects in Self-Access Center where the Internet was accessible.
3. There were 100 items for the WBCT and the scoring was one point for the exact correct answer.
4. The computer-based QPT was an adaptive test consisting of 25 items and the score and the level of each test-taker was automatically yielded on the computer based on the Look-up table for computer-based and paper and pen scores. (Table 2.1: 48). The instructor or the researcher was provided a password to get the results. The computer-based scores, which were out of 100, did not represent the number of questions answered correctly, but represented a measure of ability on the ALTE scale.
5. Before doing the WBCT, a 30-minute tutorial session and practice period were performed.
6. The Web-based C-Test was administered before the computer-based Quick Placement Test (QPT). The time allotments were 50 minutes for the WBCT and 30 minutes for the QPT.
7. The data were analyzed to obtain the means, standard deviations, the difficulty indices and the reliability of the WBCT, as well as the correlation between WBCT and QPT as shown in Tables 3.4-3.7.

The following table shows the means and standard deviations of WBCT and QPT in Pilot Study 1

Table 3.4
Means and standard deviations of WBCT and QPT
in Pilot Study 1

	Mean	Std. Deviation
WBCT	40.344	13.439
QPT (computer-based)	32.875	7.786

Table 3.4 demonstrates the means and standard deviations of the WBCT and those of the QPT in Pilot Study 1. The two sets of scores were out of 100 points. The result yields that the mean of the WBCT (40.34) is greater than that of the QPT (32.87). This indicates that the WBCT is easier than the QPT. Additionally, the standard deviation of the WBCT (13.44) demonstrates that the scores of the WBCT distributes more than those of the QPT (7.78).

The following table shows the readability, facility index, and discrimination index of the WBCT in Pilot Study 1.

Table 3.5
Readability, facility index, and discrimination index of WBCT
in Pilot Study 1

	Readability	Facility index	Discrimination index
Text1	5.6	0.438	0.448
Text2	6.4	0.406	0.508
Text3	7.0	0.344	0.438
Text4	9.4	0.408	0.377
Text5	11.3	0.430	0.338
X	7.94	0.405	0.421

According to Table 3.5, the facility indices of the 5 texts range from 0.344 to 0.438 and the discrimination indices range from 0.338 to 0.508. The average facility index is 0.405 and the average discrimination index is 0.421. This indicates that the texts implemented in the WBCT are good. The readability, which shows grade level of the whole test, is 7.94 or Grade 8. Text 1 reveals the highest facility index (0.438) while Text 3 reveals the lowest (0.344). This means that Text 1 is the easiest while Text 3 is the most difficult. In addition, Text 2 reveals the highest discrimination index (0.508) while Text 5 reveals the lowest (0.338). This means that Text 2 discriminates test-takers the best while Text 5 discriminates the worst.

The following table illustrates the reliability of the WBCT classified by subtests in Pilot Study 1.

Table 3.6

Reliability of WBCT classified by subtests in Pilot Study 1

Subtest	Items	No. of Items	Reliability
1	1-20	20	.682
2	21-40	20	.857
3	41-60	20	.605
4	61-80	20	.543
5	81-100	20	.701
Total	1-100	100	.915

According to Table 3.6, subtest 2 yields the highest reliability (.857) followed by subtest 5 (.701) while subtest 4 yields the lowest reliability (.543). The reliability of the total test is relatively high (.915). This indicates that the WBCT is reliable to assess students' proficiency.

The following table presents correlation between the WBCT and the QPT in Pilot Study 1.

Table 3.7

Correlation between WBCT and QPT in Pilot Study 1

		WBCT	QPT
WBCT	Pearson Correlation	1	.336(*)
QPT		.336(*)	1

* Correlation is significant at the 0.05 level (1-tailed). N=32

From Table 3.7, the result yields that there is a positively significant relationship between the WBCT and the QPT at the level of 0.05 and the degree of the correlation between the two tests is .336.

Pilot Study 2

Due to the limited sources of the subtests implemented in the Pilot Study 1, the committee members proposed the implementation of the text selection from a variety of sources. Therefore, 10 short reading passages from various sources were included before requesting three experts to rate the passages. The five congruence agreement texts were constructed in the WBCT and three natives were requested to do the WBCT before the implementation for the Pilot study 2.

The objective of Pilot Study 2 was to validate the five passages selected from a variety of sources. The subjects of this study were 33 first-year undergraduate students out of 670 (4.93 %) using the stratified random sampling method to obtain the sample. The subjects were from Rajamangala University of Technology Tawan-ok (RMUTTO) in Chakrabongse Bhuvanath Campus in the 2007 academic year. They were from nine majors; namely, Management, Accounting, Marketing, Economics, Information System, Computer Science, Computer Technology, English as an International Communication, and Logistics Technology as displayed in Table 3.8.

Table 3.8

Students' programs and number of students in Pilot Study 2

Students' programs	Number of students	
	Total	Pilot Study 2
1. Management	99	5
2. Accounting	43	2
3. Marketing	87	4
4. Economics	81	4
5. Information System	62	3
6. Computer Science	47	2
7. Computer Technology	79	4
8. English as an International Communication	93	5
9. Logistics Technology	79	4
Total	670	33

The following table shows the topics, word counts, readability and sources of the 10 passages in Pilot Study 2.

Table 3.9

Topics, word counts, readability and sources of the 10 passages

in Pilot Study 2

Topics	Word counts	Readability	Sources
1. A SHOCK	73	5.4	"Editor's Note". *National Geographic*. Feb 2007, Vol 211, No. 2: 6
2. DEEP BREATH-ING	95	6.4	Health, S. "The Right Breath", *Reader's Digest*. http://www.rd.com/content/openContent.do?contentId=16354
3. AUTUMN	95	6.7	http://www.teach-nology.com/worksheets/language_arts/reading_comp/elem/ver12/
4. LAUGHTER	88	7.0	"Enhance Your Sense of Humor." *Reader's Digest*. http://www.rd.com/content/openContent.do?contentId=16125
5. TAKING PICTURES	62	7.3	Icenhower, M. W. "Taking a Picture." *Ambassador Youth*. http://www.thercg.org/youth/articles/0403-tap.html
6. FISH AND WHALES	92	8.4	Gander, J. C. 2003. Fish and Whales. http://www.stemnet.nf.ca/CITE/whale_fish.pdf
7. INTELLI-GENCE PILLS	88	9.9	English Language Articles. 2007.*Intelligent pills*. http://www.usingenglish.com/comprehension/21.html
8. BARBECUE	78	10.3	Smith, A. F. editor. *2004. "Barbecue."* Oxford Encyclopedia of Food and Drink in America. Oxford University Press:New York. Volume 1: 64-65.
9. MOVIE REVIEWS	71	11.3	Reading Movie Reviews." *Bangkok Post*. http://bangkokpost.net/education/movies.htm
10. STUDY IN THE U.S	79	11.8	Barrett, M. "Is the USA Right for You?" *TOEFL access: US edition*. http://www.toeflaccess.com/articles/ETS/archive/us/

According to the table, the readability of the 10 passages ranged from 5.4 – 11.8. The passages were derived from a variety of sources.

For construct and content validations, three experts were invited to express their opinions towards the 10 short reading passages complete by themselves. There was a questionnaire for the experts to rate on the 10 prospective passages implemented in the WBCT in terms of these characteristics: general in contents, authentic either in the writers or sources, accurate in the language use, appropriate, and interesting to teenagers (first-year students). The five passages out of ten were selected and implemented online regarding the International Congruence Agreement (ICA) or the proportion of agreement among the three raters (See Appendix E: 148).

The following table illustrates the topics of five selected passages or texts with ICA average of the three experts, word counts, readability and sources of the passages.

Table 3.10

Topics of five selected passages, ICA average from three raters, word counts, readability and sources of the passages

Topics	ICA average from 3 raters	Word counts	Read-ability	Sources
1. A SHOCK	93.2	73	5.4	"Editor's Note". *National Geographic*. Feb 2007, Vol 211, No. 2: 6
2. DEEP BREATH-ING	100	95	6.4	Health, S. "The Right Breath", *Reader's Digest*. http://www.rd.com/content/openContent.do?contentId=16354
4. LAUGHTER	100	88	7.0	"Enhance Your Sense of Humor." *Reader's Digest*. http://www.rd.com/content/openContent.do?contentId=16125
7. INTELLI-GENCE PILLS	100	88	9.9	English Language Articles. 2007.*Intelligent pills*. http://www.usingenglish.com/comprehension/21.html
9. MOVIE REVIEWS	100	71	11.3	Reading Movie Reviews." *Bangkok Post*. http://bangkokpost.net/education/movies.htm

According to Table 3.10, the five selected passages were passages 1, 2, 4, 7 and 9. The ascending readability grade levels ranging from 5.4 to 11.3 were acquired

from Flesch-Kincaid Grade Level in Microsoft Word. The passages were transformed in the C-test format by the software program online for the implementation of the study. (See the five passages in full texts and in C-test format in Appendix F: 151).

Subsequently, the tests were tried out by three native speakers to verify whether the tests were appropriate for the subjects based on a basic assumption of C-tests. According to a basic assumption of the C-test, native speakers should make virtually perfect scores of 90% or higher. If native speakers cannot make scores higher than 90%, then the text should not be used with non-native speakers (Raatz and Klein-Braley, 1981). Table 3.11 shows the three natives' scores on the WBCT.

Table 3.11

Three natives' scores on the WBCT

Cases	WBCT scores of 100
Native 1	99
Native 2	95
Native 3	99
Average	97.67 %

From the table, the result shows that the native speakers' scores achieve 97.67 %. This indicates that the passages implemented in this study are appropriate for non-native speakers including the subjects of this study.

The following displays the process of the data collection and data analysis of Pilot Study 2.

1. The two tests: the Quick Placement Test (QPT) and the Web-based C-Test (WBCT) were administered to the 33 subjects. The paper and pen Quick Placement Test (QPT) was administered in the classroom instead of the computer-based QPT. This was due to limitations of some computers' capacities in the institution that could not process listening tests and could not display images. According to the Quick Placement Test Manual (2001), the computer-based and the paper and pen tests are

interchangeable. The scores of the paper and pen test can be seen in Table 2.1. They were used to classify the test-takers' levels (See Table 2.1: 48).

2. The paper and pen QPT consists of 2 parts; Part 1 (40 items) and Part 2 (20 items). The subjects of the Pilot Study 2 took both parts. The time allotment was 50 minutes. However, the subjects' scores could not meet the criteria that require the test-taker to score 36 or more to complete Part 2 of the test. So, only Part 1 was taken into account. Therefore, only Part 1 of the test taken by 33 subjects was computed to find the results.
3. The WBCT with the 5 selected passages and 100 deletions was implemented online in the Self-Access Center where the Internet was accessible. Before doing the WBCT, a 30-minute tutorial session with practice period was performed. The time allotment of the actual WBCT for Pilot Study 2 was 50 minutes.
4. The data were analyzed for the descriptive statistics, the difficulty indices and the reliability of the WBCT, as well as the correlation between WBCT and QPT as presented in Table 3.12.

Table 3.12
Means in percentage and standard deviations of WBCT and QPT in Pilot Study 2

	Mean in percentage	Std. Deviation
WBCT	37.061	13.917
QPT	40.530	9.535

Table 3.12 demonstrates the means and standard deviations of the WBCT and the QPT. Due to the difference in full scores of the two tests, the means in percentage of the tests were displayed. The result yields that the mean of the QPT (40.530) is a bit greater than that of the WBCT (37.061). This indicates that the QPT is easier than the WBCT when the QPT, Part I, was implemented. However, the standard deviation

of the WBCT (13.917) indicates that the scores of the WBCT distributes more than that of the QPT (9.535).

The following table presents the readability, facility index, and discrimination index of WBCT in Pilot Study 2.

Table 3.13
Readability, facility index and discrimination index of WBCT in Pilot Study 2

	Readability	Facility index (IF)	Discrimination index (ID)
Text1	5.4	0.246	0.336
Text2	6.4	0.426	0.414
Text3	7.0	0.380	0.368
Text4	9.9	0.344	0.441
Text5	11.3	0.458	0.327
X	8.0	0.370	0.377

Table 3.13 shows text difficulty of the WBCT. The facility indices of the 5 texts range from 0.246 to 0.458 and the discrimination indices range from 0.327 to 0.441. The average facility index is 0.370 and the average discrimination index is 0.377. This indicates that the texts implemented in the WBCT are good. The readability of the test itself is 8.0 meaning the grade level of the whole test. Text 5 reveals the highest facility index (0.458) while Text 1 reveals the lowest (0.246). This means that Text 5 is the easiest while Text 1 is the most difficult. In addition, Text 4 reveals the highest discrimination index (0.441) while Text 5 reveals the lowest (0.327). This means that Text 4 discriminates test-takers the best while Text 5 discriminate the worst.

The following table displays the reliability of the WBCT classified by subtests in Pilot Study 2.

Table 3.14

Reliability of WBCT classified by subtests in Pilot Study 2

Subtest	Items	No. of Items	Reliability
1	1-20	20	.775
2	21-40	20	.812
3	41-60	20	.804
4	61-80	20	.802
5	81-100	20	.635
Total	1-100	100	.837

From the table, Cronbach's Alpha was conducted to find the reliability of the WBCT. The result yields that the reliability of the WBCT in Pilot Study 2 is 0.837. This indicates that the test is appropriate to be used in the study. In addition, subtest 2 yields the highest reliability (.812) followed by subtest 3 (.804) while subtest 5 yields the lowest reliability (.635).

On observing the reliability of the WBCT, the test yields high reliability which indicates that the WBCT is reliable. It can be used to assess students' proficiency as it claims. More importantly, the five subtests yield moderate to high reliability. This ensures the researcher that the five subtests are reliable to implement as tests in the WBCT.

The following table illustrates the correlation between the WBCT and the QPT in Pilot Study 2.

Table 3.15

Correlation between WBCT and QPT in Pilot Study 2

		WBCT	QPT (Part I)
WBCT	Pearson Correlation	1	.293(*)
QPT (Part I)		.293(*)	1

* Correlation is significant at the 0.05 level (1-tailed). N=33

According to Table 3.15, Pearson's Product-Moment Correlation is employed to observe the correlation between the Web-based C-Test and the Quick Placement Test. The correlation of the two tests is significant at the .05 level at the degree of 0.293.

Data collection

The following steps were the data collection of the main study.

1. There were some adjustments from the two pilot studies to the main study; that is, the use of paper and pen QPT instead of the computer-based QPT; and the time allotment for WBCT's tutorial session and practice period, which was adjusted from 30 minutes to 15, and the time for the actual WBCT was adjusted from 50 minutes to 45 minutes. Therefore, before doing the WBCT, a 15-minute tutorial session and practice period necessary for doing the test was performed. And the actual WBCT was administered on line to the subjects within 45 minutes.
2. The QPT and the WBCT were administered to the 134 subjects.
3. The QPT paper and pen version Part 1 was implemented to the subjects of the study in the classroom. The time allotment for the QPT was 30 minutes.
4. The WBCT consisting of 5 selected passages with 100 deletions was administered to the subjects in the Self Access Center where the Internet was accessible. Before doing the WBCT, a 15-minute tutorial session and practice period were performed. The time allotment of the actual WBCT was 45 minutes.
5. A retrospective interview was conducted with 10 "high achievers" and 10 "low achievers" to investigate the test-takers' strategies in doing the WBCT.

Data analysis

The data collected from the main study were analyzed to answer the three research questions. The following table illustrates the research questions, statistics of data analysis, and content analysis employed to answer each research question.

Table 3.16
Research questions and the analysis

Research questions	Analysis
1. Based on the Quick Placement Test, how many levels of English proficiency can the Web-based C-Test differentiate?	Correlation and regression of the two tests were calculated on the basis that the higher the prediction, the greater the correlation. The cut-off score for each level was confirmed by the predictive equation that best fits each level.
2. What test taking strategies do the "high achievers" and "low achievers" employ in doing the Web-based C-Test?	Content analysis was performed for the qualitative data and the data were counted in numbers.
3. What are the high achievers and low achievers' opinions towards the Web based C-test?	Content analysis was performed for the qualitative data and the data were counted in numbers.

This chapter has illustrated the research procedures in terms of subject selection, research instruments, instrument development, data collection, and data analysis. The subjects in this study were randomly selected as they were from the nine different major studies. The subjects will do two tests: the QPT and the WBCT, which was constructed by means of an online software program. They will also be interviewed to investigate their test-taking strategies and opinions towards the WBCT. All the data will be analyzed and presented in the next chapter.

Along with the research procedures, a sample of the 134 subjects from the target population was selected. In the meantime, the research instruments were carefully designed to objectively obtain the criteria of the WBCT underlying the C-test procedures. During the period of instrument development, a variety of short passages were researched as to seek suitability and then narrowed down to the five selected short passages to be constructed in the WBCT. Validation of the WBCT was performed in the pilot studies. There were also some adjustments in the passages and time allotment of the administration of the two tests. As the software program was new to the researcher, knowledge of the implementation of the software program was thoroughly studied until skillful enough to employ the program for the construction, administration, and scoring of the WBCT. In other words, the researcher tried out the program by doing the tests implemented online to see how the process of the software program would proceed. In the meantime, questions and answers between the researcher and the creator of the software program were frequent to make all the issues clear before implementing the WBCT in this study.

The data collection required the cooperation of several groups of people: English instructors, first-year students' advisors, first-year students, and office workers at the Self-Access Center. Also, some documents were provided to inform the subjects with the process of doing the test together with the researcher's instructions. The accessibility to the students needed collaboration of the English teachers as well as the students' advisors, not only to inform the subjects to attend the tests, but also their permission to allow the students to take the tests at the time scheduled. The data was investigated by conducting statistical analysis of the two tests' scores and the content analysis of the information derived from the retrospective interviews. Several consultations with a statistician were made to ensure the accuracy of the interpretation of the data.

It can be perceived that there were a number of steps in the research procedures generated from this chapter, which provided constructive performances to be conducted before, during, and after the data collection and data analysis. The next chapter presents the results of the study.

CHAPTER IV

RESULTS OF THE STUDY

This chapter presents the findings of the study divided into six main parts. The first part reveals the descriptive statistics of the WBCT and the QPT. The second part deals with the text difficulty indices of the WBCT. The third part presents the reliability of the WBCT. The fourth part displays the correlation between the WBCT and the QPT. The fifth part works on the level classification of the WBCT to answer research question 1. The last part deals with the content analysis of the test-takers' strategies in doing the WBCT and their opinions towards the test to answer research questions 2 and 3.

RESULTS OF THE STUDY

Descriptive statistics of the WBCT and the QPT

The descriptive statistics of the study display the means in percentage and the standard deviations of the WBCT and the QPT derived from the calculation of the134 subjects' scores of the two tests. The means in percentage and the standard deviations of the two tests in the main study are displayed in Table 4.1.

Table 4.1
Means in percentage and standard deviations of WBCT and QPT in the Main Study

	Mean in percentage	Std. Deviation
WBCT	34.000	13.231
QPT	40.504	8.104

Table 4.1 demonstrates the means in percentage and standard deviations of the WBCT and the QPT. The mean of the QPT (40.504) is greater than that of the WBCT (34.000). This indicates that the QPT is easier than the WBCT when the QPT, Part I,

was implemented. The standard deviation of the WBCT (13.231) is greater than that of the QPT (8.104). This indicates that the scores of the WBCT distribute more than those of the QPT.

Text difficulty indices of the WBCT

The final version of the WBCT consisted of five texts making the total of 100 test items together. The text difficulty indices and readability of the WBCT in the main study were computed. The results are displayed in Table 4.2.

Table 4.2
Readability, facility index, and discrimination index of WBCT in the Main Study

	Readability	Facility index (IF)	Discrimination index (ID)
Text1	5.4	0.219	0.322
Text2	6.4	0.381	0.408
Text3	7.0	0.373	0.331
Text4	9.9	0.330	0.466
Text5	11.3	0.381	0.353
X	8.0	0.337	0.376

From the table, the facility indices of the five texts ranging from 0.219 to 0.381 indicate that the texts are not too simple or too difficult. The discrimination indices ranging from 0.322 to 0.466 indicate that the texts discriminate test-takers fairly well. The average facility index is 0.337 and the average of discrimination index is 0.376. This indicates that the WBCT is a good test. The readability of the whole test is 8.0, meaning the grade level of the WBCT is in Grade 8. Text 2 and Text 5 reveal the highest facility index (0.381) while Text 1 reveals the lowest (0.219). This means that Text 2 and Text 5 are the easiest while Text 1 is the most difficult. In addition, Text 4 reveals the highest discrimination index (0.466) while

Text 1 reveals the lowest (0.322). This means that Text 4 discriminates test-takers the best while Text 1 discriminates the worst.

Reliability of the WBCT

Reliability is one of the test qualities. The reliability of the WBCT was derived from the calculation of the scores in each subtest of 20 items and the total scores of 100 items. The reliability of the WBCT in each subtest and the total, the WBCT itself, is presented in Table 4.3.

Table 4.3

Reliability of WBCT classified by subtests in the Main Study

Subtest	Items	No. of Items	Reliability
1	1-20	20	.755
2	21-40	20	.773
3	41-60	20	.648
4	61-80	20	.816
5	81-100	20	.769
Total	1-100	100	.817

From the table, the WBCT in the main study has a reliability of 0.817 (Cronbach's alpha)[1]. This level of reliability means that the WBCT is consistent in measurement, which indicates that the test is appropriate for the study. Also, the reliability score of each subtest confirms its consistency of measurement. As perceived, subtest 4 yields the highest reliability (.816) followed by subtest 2 (.773) while subtest 3 yields the lowest reliability (.648).

Correlation between the WBCT and the QPT

The correlation between the WBCT and the QPT was preliminarily investigated so as to observe the concurrent validity of the two tests. Pearson's Product-Moment Correlation was employed to compute the correlation between the

[1] All computations were carried out with SPSS version 10.

WBCT and the QPT. The result yields that the correlation of the two tests is significant at the .01 level at the degree of 0.335 as displayed in Table 4.4.

Table 4.4
Correlation between WBCT and QPT in the Main Study

		WBCT	*QPT (Part I)*
WBCT	Pearson Correlation	1	.335(**)
QPT (Part I)		.335(**)	1
	Sig. (2-tailed)	.000	.

** Correlation is significant at the 0.01 level (2-tailed).

To correlate the two sets of scores on the same scale, normalized T-scores were performed. To elaborate, the full score of the WBCT, which was 100, and that of the QPT, which was 40, were converted into normalized T-scores to locate the two sets of scores into the same scale before investigating their correlation (See Appendix G: 156 and Appendix H: 157). The result yields that the correlation of the two tests is significant at the .01 level at the degree of 0.340 which is a bit higher than the first calculation when the scores are in different scales. This level of the relation means that the concurrent validation of the two tests is limited. The correlation between the WBCT and the QPT is exhibited in Table 4.5.

Table 4.5
Correlation between WBCT and QPT in normalized T-scores
in the Main Study

		WBCT	*QPT(Part I)*
WBCT	Pearson Correlation	1	.340(**)
QPT(Part I)		.340(**)	1
	Sig. (2-tailed)	.000	.

** Correlation is significant at the 0.01 level (2-tailed).

Level classification of the WBCT

The purpose of the WBCT is to place students into appropriate levels of language ability; thus, its potential in performing such a task is relatively important. To serve that purpose, the cut-off score for each level was examined using the Association of Language Testers in Europe (ALTE) levels. This is in accordance with the QPT which reports test results as a band on the ALTE. In this case, the QPT's three low levels were explored after the WBCT's scores of 100 points were converted to 40 points (See Appendix I: 159). According to the Look-up table for computer-based and paper and pen scores, Part I using scores of 40 (See Table 2.1: 48) was employed in the WBCT's level classification. The results of the level classification of the WBCT addressed in the first research question yield that the scores of the 134 subjects are classified into three different levels; that is, Levels 0-2. There are ninety (90) subjects at Level 0, forty-one (41) subjects at Level 1, and three (3) subjects at Level 2 as displayed in Table 4.6.

Table 4.6

WBCT levels and number of the subjects

WBCT Levels	No. of subjects
Level 0	90
Level 1	41
Level 2	3
Total	134

To test the hypothesis that there are significant differences among the mean scores of the subjects at each level, Level 2, which contained only 3 subjects, was treated as an outlier. Therefore, the two levels, Level 0 and Level 1 were taken into computation employing Independent Sample t-test to test their mean difference. The result yields that there is a significant difference between Level 0 and Level 1 at the .01 level. The means, standard deviations and t value of the two levels are displayed in Table 4.7.

Table 4.7

Means, standard deviations and t value of Level 0 and Level 1

Level	N	Mean	Std. Deviation	Std. Error Mean	t value
0	90	26.91	9.11	.96	15.497 *
1	41	47.37	5.80	.91	

* $p \leq .01$

To address the first research question, based on the Quick Placement Test, WBCT can differentiate English proficiency of the test-takers into two levels; that is, Level 0 to Level 1. According to ALTE Levels (See Table 2.1: 48), Level 0 refers to 'Beginner', and Level 1 refers to 'Elementary.' The features of ALTE levels show that Level 0 or Beginner refers to Breakthrough Level and it features the test-taker's basic ability to communicate and exchange information. Level 1 or Elementary refers to Level 1 (Waystage) and it features the test-taker's capacity to deal with simple straightforward information and to get by in familiar contexts (See Table 2.2: 50).

Furthermore, a test of regression was employed to investigate the degree of prediction of the WBCT compared to the QPT so as to obtain a formula to predict scores in the QPT. As a result, the two tests' T-scores were computed on the basis that the greater the prediction, the higher the correlation as displayed in Table 4.8.

Table 4.8

Test of regression of WBCT and QPT

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.340	.116	.109	9.243

a. Predictors: (Constant), WBCT

According to the table, the result yields that the correlation of the two tests is significant at the .01 level with the degree of relation of .340. R square value equals .116. In order to illustrate the regression equation, the beta and constant values obtained from the analysis are displayed in Table 4.9.

Table 4.9

Beta and constant value of WBCT and QPT

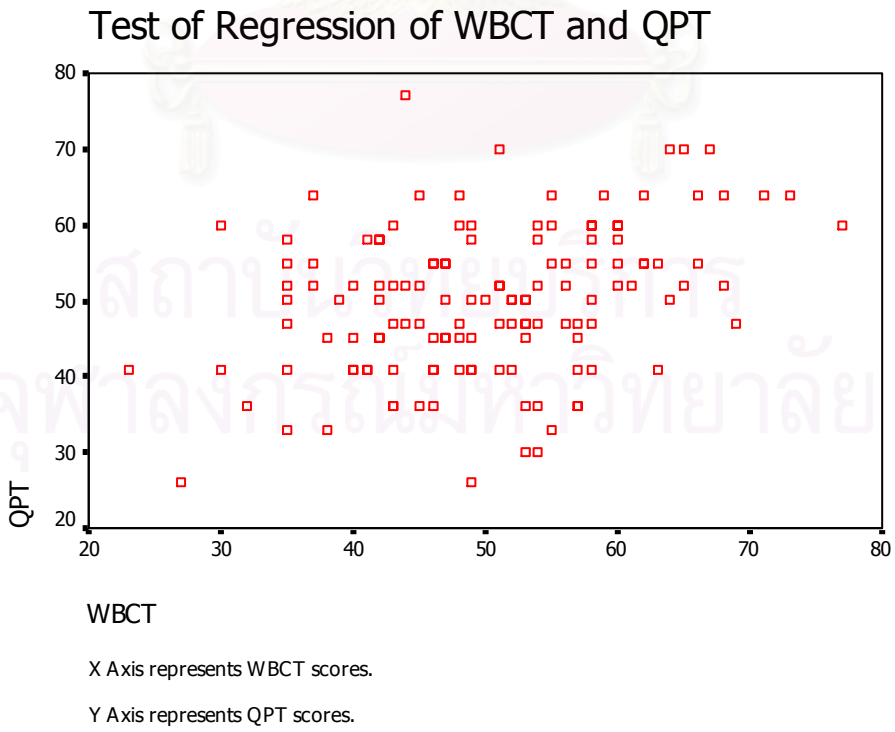
Model		Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	Sig.
		B		Beta		
1	(Constant)	33.393	4.069		8.207	.000
	WBCT	.331	.080	.340	4.157	.000

a. Dependent Variable: QPT

From the table, the regression equation is y' = 33.393+.331x when y' is the predicted score in the QPT and x is a score in the WBCT. The equation can be applied to predict the score in the QPT when the score in the WBCT of each subject is available.

In Figure 3.1, the Scatterplot graph displays the test of regression between the WBCT and the QPT; where X Axis represents scores of WBCT and Y Axis represents scores of QPT.

Figure 3.1



It is noticeable that there is no linearity between the scores of the two tests due to the limited correlation between the WBCT and the QPT (.340). This results in the minimal R square value (.116) leading the predictor variable (WBCT) to account for 11.6 % of the variance in QPT's scores. It shows that the predictive efficiency of WBCT's scores to QPT's scores is limited.

Content analysis

To address the second and third research questions a retrospective interview was performed with 10 "high achievers" and 10 "low achievers" to investigate their test-taking strategies in doing the WBCT and to explore their opinions towards the developed test. The subjects of this study were classified into 2 levels by the WBCT. Thus, the "low achievers" referred to representatives at Level 0 and the "high achievers" referred to representatives at Level 1. In other words, 10 subjects from Level 0 as "low achievers" and 10 subjects from Level 1 as "high achievers" were interviewed by the researcher. The subjects for the retrospective interview were randomly selected to observe their test-taking strategies and their opinions towards the WBCT.

To address the second research question on "What test-taking strategies do the "high achievers" and "low achievers" employ in doing the WBCT?", the "high achievers" and the "low achievers" were asked 5 questions based on the WBCT-taking strategies compiled by the researcher (See pp. 56-57). The results are displayed in Tables 4.10-4.14. The number indicates the subjects in each category. The following table illustrates the number of high and low achievers' test-taking strategies acquired from Question I.

Table 4.10

High and low achievers' test-taking strategies (Question I)

Questions	"High achievers"	"Low achievers"
I. What were your strategies in doing the test?		
1. Reading the clause	7	5
2. Reading the sentence	8	4
3. Scanning the whole text	5	2
4. Reading each word or parts of sentences	4	6
5. Using extra knowledge	4	2
6. Using background knowledge	6	3
7. Guessing	6	10
8. Finding answer in the passage	6	2
9. Analyzing the passage	2	1
10. Translating words or sentences in the context	8	5
11. Using grammar knowledge	5	3
12. Others:		
12.1 Highlighting the missing parts before restoring	2	0
12.2 Using the space bar to check a number of the missing letters	1	0
12.3 Restoring first the words that the test-takers are confident about	1	1
12.4 Restoring first the familiar words	0	1
12.5 Restoring short words	0	1
12.6 Looking at the preceding words	3	1
12.7 Thinking of the main idea of the passages	1	0
12.8 Counting the missing letters	0	1

To address Question 1, "What were your strategies in doing the test?", it is remarkable that "high achievers" and "low achievers" used different strategies in doing the WBCT. To elaborate, the "high achievers" seemed to agree with most of

the WBCT-taking strategies, except for items 4 and 7, which the “low achievers” employed most; that is, reading each word or parts of sentences (6) and guessing (10). The “high achievers” used item 2 and item 10 most, reading the sentence (8) and translating words or sentences in the context (8). There were other interesting strategies, which the “high achievers” implemented in doing the WBCT; namely, looking at the preceding words (3), highlighting the missing parts (2), using the space bar to check the number of missing letters (1), restoring first the words that the test-takers are confident about (1), and thinking of the main ideas of the passages (1). The “low achievers” also presented other strategies in doing the test; they were: restoring first the words that the test-takers are confident about (1), restoring first the familiar words (1), restoring short words (1), looking at the preceding words (1), and counting the missing letters (1). Interestingly, among the supplementary strategies, there were only two additional strategies involved using the computer (12.1 and 12.2). The others indicated strategies in restoring the passages.

The following table shows the high and low achievers’ test-taking strategies acquired from Question II.

Table 4.11

High and low achievers' test-taking strategies (Question II)

Questions	"High achievers"	"Low achievers"
II. What in the passage helped you restore the words most?		
a) The title of the passage	8	2
2. The first sentence (left standing)	7	3
3. The first half of the word.	9	5
4. The rest of the text	1	0
5. The word restored	5	2
6. Others:		
6.1 The context in the passages	3	0
6.2 The repeated words	2	1
6.3 The other sentences in the passages	1	0

As for Question 2: "What in the passage helped you restore the words most?", the "high achievers" responded that the first half of the word (9) and the title of the passages (8) helped them restore the words most while the "low achievers" answered that the first half of the word (5) and the first sentence left standing (3) assisted them in restoring the missing parts. There were also other things that facilitated their restoring the missing parts. For "high achievers", the context in the passages (3), the repeated words (2), and the other sentences in the passages assisted them to restore the letters while the "low achievers" expressed that the repeated words (1) assisted the test-taker in restoring the letters.

The following table presents the high and low achievers' test-taking strategies acquired from Question III.

Table 4.12

High and low achievers' test-taking strategies (Question III)

Questions	"High achievers"	"Low achievers"
III. What kinds of words were most difficult to restore?		
1. 2-3 letters (short words)	2	0
2. 4-6 letters. (medium length words)	5	6
3. more than 6 letters (long words)	4	8
4. Others:		
4.1 Unseen words	2	5
4.2 Unable-to-spell words	1	4
4.3 Forgotten words	0	1

For Question 3: "What kinds of words were most difficult to restore?", the most difficult words for the "high achievers" were medium length words with 4-6 letters (5) while long words with more than 6 letters were difficult for the "low achievers" (8). Interestingly, the "high achievers" added that the unseen words (2) and unable-to-spell words were the most difficult words while the "low achievers" expressed that the unseen words (5), the unable-to-spell words (4), and the forgotten words (1) were the most difficult words. The following table illustrates the interviewees' responses concerning test-taking strategies acquired from Question IV.

Table 4.13

High and low achievers' test-taking strategies (Question IV)

Questions	"High achievers"	"Low achievers"
IV. What knowledge most helped you restore the words?		
1. Vocabulary.	8	10
2. Grammar	5	2
3. Reading comprehension	7	3
4. General proficiency	8	4
5. Others: …….		

When asked Question 4: "What knowledge most helped you restore the words?", the "high achievers" asserted that they implemented vocabulary (8), and general proficiency (8) to help in restoring the words. The "low achievers" expressed similarly; however, they were different in numbers. That is, the "low achievers" employed vocabulary (10) and general proficiency (5) in doing the test. The following table shows the high and low achievers' test-taking strategies acquired from Question V.

Table 4.14

High and low achievers' test-taking strategies (Question V)

Questions	"High achievers"	"Low achievers"
V. Did you understand all the five subtests after restoring the words?		
1. Yes, most of them. (3-5)	6	2
2. Yes, a few. (1-2)	4	6
3. Not at all.	0	2
4. Others……		

For Question 5: “Did you understand all the five subtests after restoring the words?”, the “high achievers” agreed that they understood three to five passages (6), while the “low achievers” expressed that they comprehended one to two passages (6).

It can be concluded that there were many similarities and differences between the “high achievers” and the “low achievers” in implementing test-taking strategies in doing the WBCT. Eight “high achievers” employed reading the sentence and translating words or sentences in the context while ten “low achievers” used guessing and six used reading each word or parts of sentences. Additionally, the “high and low achievers” made use of ‘the first half of the words’ most, but in different degrees (9 and 5 respectively). Besides, five “high achievers” expressed that medium length words with 4-6 letters were difficult to restore while eight “low achievers” stated that long words with more than 6 letters were difficult to restore. Moreover, eight “high achievers” employed vocabulary and general proficiency in restoring the words while ten “low achievers” employed vocabulary in restoring the words. When asked if they understood the passages after restoring the words, six “high achievers” expressed that they understood three to five passages, while six “low achievers” expressed that they understood one to two passages. Finally, both groups included other interesting strategies in doing the test.

To answer the third research question on “What are the opinions of the “high achievers” and the “low achievers” towards the WBCT?”, the “high achievers” and the “low achievers” were asked 5 questions. Tables 4.15-4.17 show the high achievers’ and the low achievers’ opinions towards the WBCT.

The following table illustrates the high achievers’ and the low achievers’ opinions towards the WBCT acquired from Questions 1-3.

Table 4.15
High and low achievers' opinions towards the WBCT
(Questions 1-3)

Questions	"High achievers"	"Low achievers"
I. What do you think the test tests?		
1. General English ability	6	4
2. Vocabulary	2	6
3. Reading comprehension	1	0
4. Grammar	1	0
5. Others: ……		
II. Do you like doing the test on the Web-based delivery?		
1. Yes.	10	9
2. No.	0	1
3. Others: …….		
III. Do you think the test can assess your English proficiency?		
1. Yes.	9	7
2. No.	0	0
3. Others:		
3.1 No, the test assesses vocabulary.	0	3
3.2 Not all, but some proficiency.	1	0

There were five questions to inquire their opinions in doing the WBCT. Question 1 was employed to observe what the test assessed. The majority of "high achievers" (6) expressed that the test assessed their general proficiency while the majority of "low achievers" (6) believed that the test evaluated their vocabulary and four thought that the test assessed their general proficiency. For Question 2, which inquired test-takers' opinions on the Web-based test delivery, both "high achievers" and "low achievers" expressed their preference for the Web-based test delivery (10, 9 respectively). The subsequent Question 3 observed whether the test could assess their proficiency. Nine "high achievers" out of ten agreed that the test could assess their

proficiency. As for the "low achievers", seven achievers agreed that the test could assess their proficiency while three achievers thought that the test tested vocabulary.

The following table presents the high achievers' and low achievers' opinions towards the WBCT acquired from Question 4.

Table 4.16
High and low achievers' opinions towards the WBCT (Question 4)

Questions	"High achievers"	"Low achievers"
IV. What is your opinion towards the test?		
1. Challenging.	9	9
2. Difficult.	5	8
3. Easy.	2	0
4. Boring.	0	1
5. Others:		
5.1 The test is interesting.	8	6
5.2 The feedback scores displayed on screen after submission of each test are motivating.	1	1
5.3 The feedback scores make the test interesting.	4	2
5.4 The test will be enjoyable if test-takers can restore the words.	1	2
5.5 The test is like a game.	1	0
5.6 The first half causes confusion to restore.	1	1
5.7 The letters that the test-takers fill in do not fit the blanks.	1	0

Question 4 asked test-takers' opinions towards the WBCT. Nine "high achievers" stated that the test was challenging and five thought that it was difficult. Eight "high achievers" also added that the test was interesting and four believed that

the feedback scores displayed after submission of each test were motivating. Only one high achiever thought that the feedback scores made the test interesting; the test was like a game; the first half caused confusion to restore; and the letters that the test-takers filled in did not fit the blanks. Similarly, nine "low achievers" expressed that the test was challenging, and eight believed that it was difficult. Six "low achievers" also included that the test was interesting. Two "low achievers" thought that the feedback scores made the test interesting and the test would be enjoyable if test-takers could restore the words. Only one said that the first half caused confusion to restore.

The following table illustrates the high achievers' and the low achievers' opinions towards the WBCT acquired from Question 5.

Table 4.17
High and low achievers' opinions towards the WBCT (Question 5)

Questions	"High achievers"	"Low achievers"
V. What do you think about the time allotment?		
1. Sufficient.	9	10
2. Too short.	1	0
3. Too long.	0	0
4. Others …….		

The last question was about time allotment. Both high and "low achievers" replied that the time allotment was sufficient (9 and 10 respectively). Only one high-achiever expressed that the time allotment was too short.

It can be concluded that there were also many similarities and differences between high and low achievers' opinions towards the WBCT. In regard to similarities, both high and "low achievers" expressed that the WBCT tested general proficiency and they liked the Web-based test delivery. Also, they agreed that the WBCT could assess their English proficiency. Moreover, they thought that the test was challenging and difficult. Additionally, they added that the test was interesting

and the time allotment was sufficient. In terms of differences, the majority of the "high achievers" accepted that the test tested their general proficiency while the majority of the "low achievers" expressed that the test tested vocabulary.

In conclusion, this study can answer the three research questions posted in Chapter I. First, the Web-based C-Test (WBCT) can differentiate the students into two levels: Level 0 or "Beginner" (a basic ability to communicate and exchange information) and Level 1 or "Elementary" (a capacity to deal with simple straightforward information and to get by in familiar contexts). Secondly, the high and "low achievers" implemented different strategies in doing the Web-based C-Test. The "high achievers" employed 1) reading the sentence and 2) translating words or sentences in the context while the "low achievers" used 1) guessing and 2) reading each word or parts of sentences. Moreover, the "high and low achievers" made use of the first half of the words. The "high achievers" also employed vocabulary and general proficiency in restoring the words while the "low achievers" employed vocabulary in restoring the words. Finally, the high and "low achievers" had different opinions towards the WBCT. The majority of the "high achievers" perceived that the WBCT assessed general proficiency while the "low achievers" perceived that the test tested vocabulary.

CHAPTER V

SUMMARY, DISCUSSION, CONCLUSION, AND RECOMMENDATIONS

This final chapter summarizes the findings of the study together with the discussion, conclusion, and recommendations. It aims to provide the overall picture of this study in brief as well as suggestions for language assessment.

The general characteristics of the test are summarized in terms of descriptive statistics, reliability of the WBCT, and the correlation of the WBCT and QPT so as to provide fundamental information of the WBCT.

Descriptive statistics

The mean of the QPT (40.504) is greater than that of the WBCT (34.000). This indicates that the QPT (Part I) is easier than the WBCT. However, the standard deviation of the WBCT (13.231) indicates that the scores of the WBCT distribute more than those of the QPT (8.104).

Reliability of the WBCT

The Web-based C-Test has a reliability of .817. This level of reliability means that the WBCT is consistent in measurement, which indicates that the test is appropriate for the study.

Correlation between the WBCT and the QPT

The two sets of scores were transformed to the same scale with normalized T-scores prior to finding the correlation between the two tests. The correlation between the two tests yields significantly at the .01 level at the degree of 0.340, which is a bit higher than the first calculation where the scores are in different scales.

SUMMARY OF THE FINDINGS

The main findings of this study are presented threefold, namely: Level classification of the WBCT, the test-takers' strategies in doing the WBCT, and the opinion of the test-takers towards the WBCT.

Level classification of the WBCT

The scores of the 134 subjects were classified into three different levels; namely, Levels 0-2. There are ninety (90) subjects at Level 0, forty-one (41) subjects at Level 1, and three (3) subjects at Level 2. However, Level 2, which contains only three subjects, was treated as an outlier. Subsequently, the two levels, Level 0 and Level 1 were taken into computation to test their mean differences. The results yield that there is a significant difference between Level 0 and Level 1 at the .01 level.

Therefore, to address the first research question, based on the Quick Placement Test, WBCT can differentiate English proficiency of the test-takers into two levels; that is, Level 0 and Level 1. The two levels are significantly different at the .01 level. According to ALTE Levels (See Table 2.1: 48), Level 0 refers to 'Beginner', and Level 1 refers to 'Elementary.' At Level 0 or 'Beginner,' ALTE features a basic ability to communicate and exchange information. At Level 1 or "Elementary," it features the test-taker's capacity to deal with simple straightforward information and to get by in familiar contexts (See Table 2.2: 50).

Likewise, to investigate the degree of prediction, a test of regression of the two tests in T-scores was computed. The result yields that the correlation of the two tests is significant at the .01 level. The degree of correlation is .340. The R-square value equals .116; therefore, the predictor variable (WBCT) accounts for 11.6 % of the variance in QPT's scores. It shows that the predictive efficiency of WBCT's scores to QPT's scores is limited.

Test-takers' strategies in doing the WBCT

To address the second research questions "What test-taking strategies do the "high achievers" and "low achievers" employ in doing the WBCT?", a retrospective interview was performed with 10 "high achievers" and 10 "low achievers" to

investigate their test-taking strategies in doing the WBCT. It can be concluded that there were many similarities and differences between the "high achievers" and the "low achievers" in implementing test-taking strategies in doing the WBCT. The "high achievers" employed 1) reading the sentence and 2) translating words or sentences in the context while the "low achievers" used 1) guessing and 2) reading each word or parts of sentences. Additionally, the high and "low achievers" made use of 'the first half of the words.' Moreover, the "high achievers" employed vocabulary and general proficiency in restoring the words while the "low achievers" employed only vocabulary in restoring the words.

Test-takers' opinions towards the WBCT

To address the third research question, "What are the opinions of the 'high achievers' and the 'low achievers' towards the WBCT?", a retrospective interview was performed with 10 "high achievers" and 10 "low achievers." It can be concluded that there were also many similarities and differences between the high and low achievers' opinions towards the WBCT. In regard to similarities, both high and "low achievers" expressed that the WBCT tested general proficiency and that they liked the Web-based test delivery. Also, they agreed that the WBCT could assess their English proficiency and that the test was interesting, challenging and difficult. In terms of differences, the majority of the "high achievers" accepted that the test tested their general proficiency while the majority of the "low achievers" expressed that the test tested vocabulary.

DISCUSSION

The following section presents the discussion based on the three research questions mentioned in Chapter I.

Research question 1:

Based on the Quick Placement Test, how many levels of English proficiency can the Web-based C-Test (WBCT) differentiate?

The first research question has revealed that based on the Quick Placement Test the WBCT can differentiate English proficiency of the test-takers into two levels: Level 0 or Beginner and Level 1 or Elementary. It is likely that the WBCT can differentiate the subjects into two low levels because of the subjects' limited proficiency or their low-level processing in their proficiency of the language. According to the research studies, C-tests tend to differentiate high-level processing students (Sigott, 2002; Hastings, 2002b). It is possible that the subjects in this study have limited proficiency that makes the level classification limited. This is in accordance to Rouhani's study (2008), which found that the C-test had difficulty discriminating between participants of lower and upper intermediate levels. However, there are research studies that support the C-test's discrimination power among ESL learners. Ikeguchi (1998) found that the ESL students in Japan can be differentiated into levels by the two types of C-tests implemented in the study. Unlike ESL students, the subjects in this study were EFL learners, whose proficiency may not be compatible to ESL learners. This may make the level classification restricted.

Despite the two low levels found in this study, which were due to the limitations of the subjects' proficiency, this information is useful for the institution. According to ALTE Levels, Level 0 or "Beginner" features a basic ability to communicate and exchange information, and Level 1 or "Elementary" features a capacity to deal with simple straightforward information and to get by in familiar contexts (See Table 2.2: 50). This information is, therefore, useful for the institution not only to place students in appropriate levels but also to organize English courses for them. The features of each level illustrate the capacities of the students. When students are placed in appropriate levels, both teaching and learning conditions are greatly enhanced.

As for the correlation between the WBCT and the QPT, it yields significantly at .340, which indicates that the correlation between the two tests is minimal. This can be explained that despite the two tests' purposes of measuring general proficiency, the two tests differ in various aspects. The QPT, like most tests, measures certain skills in separate parts; for example, reading, listening and writing

while the WBCT, implementing the C-test format, measures integrative skills. The WBCT is not divided into parts to measure certain skills. Instead, it presents the test in five short passages. Moreover, the second half of the words contains semantic, syntactic, morphological, lexical, and orthographic processing (Hastings, 2002b). Therefore, test-takers are to integrate contextual information while doing the C-test. In addition, the two tests differ in test formats in that the QPT is a multiple choice test while the WBCT requires test-takers to fill in the missing characters. This may affect the correlation between the two tests. In this study, it can be concluded that the WBCT does not entirely measure the same type of ability as the QPT despite the similar purpose of measuring general proficiency.

The finding indicates that there is a significant difference between Level 0 and Level 1 at the .01 level. As a result, Hypothesis 1: "There are significant differences among the mean scores of the students at each level" is achieved. This means that the WBCT can differentiate the test-takers according to their proficiency.

Research question 2:

What are the test-taking strategies that the "high achievers" and "low achievers" employ in doing the Web-based C-Test?

In search of test-takers' strategies, the WBCT-taking strategies was developed in this study focusing on the reading strategies and cloze testing strategies. To elaborate, the WBCT-taking strategies was derived from a combination of Bachman's test-taking framework (1985), Cohen's micro strategies (1989), Nevo's test-taking strategies (1989 cited in Cohen, 1994: 126), and Hosenfeld's reading strategies (1977 cited in Cohen, 1994: 123). The test-takers' strategies were elicited from a retrospective interview. Interestingly, the "high achievers" responded to most of the WBCT-taking strategies in doing the test more than the "low achievers" except the two items: item 4 and item 7, reading each word or parts of sentences and guessing, which the "low achievers" responded to both items most (See Table 4.10: 88).

According to the interviews, the "high achievers" employed 1) reading the sentence and 2) translating words or sentences in the context most while the "low

achievers" used 1) guessing and 2) reading each word or parts of sentences most. It can be perceived that the "high and low achievers" employed strategies pertinent to Cohen's micro-level strategies (1994: 142), which consist of two levels: word-level and sentence-level. The "high achievers" were likely to implement sentence-level strategies while the "low achievers" were likely to implement word-level strategies. Additionally, both "high and low achievers" employed reading strategies in doing the WBCT. The "high achievers" employed translating words or a sentence in the context, which is pertinent to Hosenfeld's reading strategies (1977 cited in Cohen, 1994: 123), while the "low achievers" employed guessing, which is pertinent to Nevo's reading strategies (1989 cited in Cohen, 1994: 126) stating that the guessing is not based on any rationale.

In addition, the "high and low achievers" made use of the first half of the word (See Table 4.11: 90). This can be explained by the reduced redundancy theory of C-tests. This concept starts with the assumption that educated adult native speakers of a language can generally make use of the redundancy of their language to restore damaged messages. According to the two initiators, Klein-Braley and Raatz (1984), a redundant message contains more information than is essential for the understanding of the message. Therefore, if parts of the message are damaged or lost, it can be restored from those parts that have been left intact. This indicates the practice that both "high and low achievers" inevitably followed this concept of reduced redundancy that helped them to restore the missing parts.

Moreover, the "high achievers" expressed that they employed vocabulary and general proficiency in restoring the words. This is according to the C-test format, the second half deletion occurs in both content words and function words. The test-takers tend to employ their knowledge of semantic, syntactic, morphological, lexical, and orthographic processing to fill in the missing parts (Hastings, 2002b). Therefore, it was likely that the "high achievers" employed general proficiency including their knowledge of vocabulary to restore the missing parts. In doing so, the WBCT could assess their proficiency of the language. Unlike the "high achievers", the "low achievers" stated that they employed only vocabulary in restoring the missing parts. It

is possible that the "low achievers" had less language ability and most low achievers thought that they had a limited repertoire when they could not restore a word. According to this testing method, it is not the redundancy of the text that is measured, but the examinee's ability to make use of the general redundancy of the language to restore the damaged text (Raatz and Klein-Braley, 1981). It can be seen that learning test-takers' strategies provides useful information for teachers and researchers to perceive how the test-taker answers the test as Cohen (1998: 37) suggests that "it is significant not only to examine learners' final products or the results from testing but also the process by which they think, choose, and produce the appropriate answers in testing situation." It is likely that different test-takers always perform differently because they differ in their strategies to solve problems.

The finding confirms hypothesis 2 that students at the high and low levels use different strategies in doing the WBCT.

Research question 3:

What are the high achievers and low achievers' opinions towards the Web-based C-Test?

The "high and low" achievers had different opinions in what the WBCT assessed. The "high achievers" expressed that the WBCT tested general proficiency while the "low achievers" expressed that the test tested vocabulary. The opinions of the "high achievers" corresponded well with the WBCT quality in that it does not assess only vocabulary but also general proficiency. According to Klein-Braley and Grotjahn (1995 cited in Raatz & Klein-Braley, 2002), C-tests do not reveal small increments in learning progress as much as they do general proficiency. Also, Dörnyei and Katona (1992, 1993) indicated that the C-test correlated well with vocabulary and grammar scores as well as speaking skills. Additionally, the second-half word deletion happens on both content and function words, therefore, in order to fill in the missing parts, the test-takers are required to employ their knowledge of semantic, syntactic, morphological, lexical, and orthographic processing (Hastings, 2002b). It is probable that the "low achievers" may have been less proficient in providing integrative knowledge to resolve the missing parts and that most "low

achievers" believe that they could not do the tests because they had not enough vocabulary.

The finding supports hypothesis 3 that "high and low achievers" have different opinions towards the Web-based C-Test in terms of what the WBCT tests.

Lastly, when addressing their preference for the WBCT, the majority of both "high and low achievers" yielded contented satisfaction towards the WBCT in that the test was interesting and challenging and that it could assess their proficiency. This can be assumed to refer to both, the C-test format and the Web-based delivery.

In terms of the C-test format, where every second half of every second word is deleted leaving the first half standing, the first half must have attracted the test-takers to restore the mutilated parts because most "high and low achievers" expressed that the first halves helped in restoring the missing parts. Only one of the "high achievers" and one of the "low achievers" expressed that the first half left standing caused confusion (See Table 4.16: 95). This can be explained by the strategic "Rule of 2" that the deletion of every second word promotes the representativeness of the sample and the half word deletion makes the test relatively simple to restore. Also, there is enough redundancy in the first half for a proficient person to reconstruct the missing part correctly (Hastings, 2002a). It is probable that those who expressed confusion on the first halves had little knowledge background of the language, such as semantic and syntactic features, which are required to restore the mutilated characters.

In terms of the Web-based delivery, the "high and low achievers" expressed their preference for the feedback scores yielded on the screen (See Table 4.16: 95) which is one of the characteristics of the Web-based delivery that can interactively yield the test results. This is pertinent to the advantages of the Web-based delivery stating that the scoring is automatic and immediate and the Web-based tests can be displayed regardless of place or time (Röver, 2001a). In addition, the test-takers' preferences for the Web-based delivery may derive from the advantages of the computer-based tests compared to the paper-based tests. It is obvious that the computer-based tests relate to the Web-based tests in that the Web-based tests are developed from the computer-based tests and they utilize computers as a test delivery medium. Kantahoe et al. (2003) proposed a computer-based S-Test to cope with the

problems that occurred in the paper-based test and the result yielded that the computer-based test could promote time and place performances. It is probable that the test-takers' preferences for the WBCT related to their preferences for taking the WBCT on the computer rather than on the paper and pencil base.

Also, there are other variables that reflect the test-takers' preferences according to Kiratibodee's study (2006) on the relationships between selected variables and reading comprehension using Computer-Based Test (CBT) performance. The findings revealed that computer familiarity was a significant predictor of CBT scores for students of average language ability, while computer attitudes were a significant predictor of CBT scores for students with high language ability. It can be assumed that average or proficient test-takers who were familiar with computers and had positive computer attitudes were likely to express preferences for the WBCT, which utilized computers as the medium of the test. However, it is probable that also "low achievers" who expressed their preference for the WBCT might have computer familiarity and positive computer attitudes.

It can be concluded that the students at "high" and "low" levels have different opinions towards the Web-based C-Test in terms of WBCT measurement. Additionally, the majority of both "high and low" achievers expressed their preference for the WBCT.

CONCLUSION

This study has answered the three research questions stated in the first chapter. First, the WBCT can differentiate the test-takers into 2 levels: Level 0 or Beginner and Level 1 or Elementary. Second, students at the high and low levels use different strategies in doing the WBCT in that the "high achievers" employed 1) reading the sentence and 2) translating words or sentences in the context most, while the "low achievers" used 1) guessing and 2) reading each word or parts of sentences most. Third, the students at "high" and "low" levels have different opinions towards the WBCT in terms of measurement. However, the majority of both "high and low" achievers expressed their preference for the WBCT.

It can be concluded that the WBCT may be implemented as an English Placement Test for Rajamangala University of Technology Tawan-ok, Chakrabongse Bhuvanath Campus. Moreover, the investigation of test-takers' strategies in doing the WBCT provided information of how students employed their knowledge in doing the test. This can lead to improvements in language learning and teaching. Additionally, the test-takers' preferences for the WBCT give light to the possibility of implementing the C-test format with the Web-based delivery for the extensive use of the WBCT. Expectedly, the WBCT can be adapted and used as a placement test for Thai students at the tertiary level in other institutions. Lastly, this inaugural study of the WBCT in Thai context may encourage other language teachers to construct their own Web-based C-Tests as an alternate measure for testing the language skills of Thai learners.

RECOMMENDATIONS

The recommendations from this study are presented threefold: Recommendations for implementations of the WBCT, recommendations for pedagogical implications, and recommendations for future research studies.

Recommendations for implementations of the WBCT

There are two stages for the implementations of the WBCT: 1) The immediate implementation and 2) the research studies for the WBCT text bank.

1. The immediate implementation

According to the level classification, the WBCT can differentiate the subjects into two ATLE levels. This information can be used in decision making for the institution to organize courses for the first-year students before taking compulsory English courses.

In practice, when the institution implements the WBCT as an English placement test, all first-year students are required to take the test and they will be classified into certain levels according to their proficiency. It is advisable that students, who are classified at Level 0 or Beginner, are required to take a Basic English Course before taking the compulsory courses. This is because the feature of Level 0 presents that the

student at this level has a basic ability to communicate and exchange information, which may not be enough to succeed in learning English at a tertiary level. The institution should take this issue into consideration and should consider a preparation course to fundamentally help them acquire additional knowledge before taking courses. The organization of this special course should be collaborated among all affected English instructors.

By contrast, students who are classified at Level 1 or Elementary can take compulsory English courses. This is because the feature of Level 1 presents that a student at this level has a capacity to deal with simple straightforward information and can get by in familiar contexts, which may more or less exhibit some skills to enter a compulsory course.

It is obvious that the level classification of the WBCT assists the placement decision which is, according to Hughes (2003), at low-stake. If students are misplaced by the test, they can usually be moved to a more appropriate class.

At a positive viewpoint, students who are placed at Level 0 benefit from the placement decision in that the institution takes great concern to provide them with courses to help them succeed in learning while students who are placed at Level 1 are grouped appropriately to facilitate teaching and learning. This agrees with Brown's idea (1996) stating that placement decisions help teachers to assign students to their ability levels and appropriate placement can enhance the teaching and learning process.

2. The research studies for the WBCT text bank

In conducting the research studies for the WBCT text bank, there are two guidelines for the institution to take into consideration: the guideline for WBCT research studies and the guideline for the implementation of the WBCT as a placement test.

a. The guideline for WBCT research studies

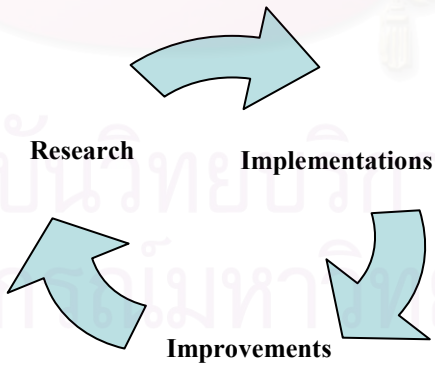
The guideline for WBCT research studies aims at conducting more subtests for the WBCT to be implemented as an English placement test for the institution so that a number of tests will be stored in the WBCT text bank to be implemented

interchangeably to prevent dishonesty in doing the test. The guideline can be initially perceived as follows:

1) A project of the construction of the WBCT as a placement test for the institution will be proposed to the institution's administrators and English instructors together with the results of this research study.
2) Discussion for the construction of the WBCT among English instructors will occur.
3) Research studies on the text selection for the WBCT regarding the reliability and text difficulty will be conducted every semester so that more subtests can be implemented in the WBCT.
4) The WBCT will be implemented as an English placement test for the institution.
5) Improvements on the WBCT will occur on an on-going basis. The on-going process for WBCT research studies is displayed as follows:

Figure 5.1

On-going process for WBCT research studies



b. The guideline for the implementation of the WBCT as a placement test

The guideline for the implementation of the WBCT as a placement test aims at placing students into appropriate levels pertinent to their proficiency. The guideline can be primarily viewed as follows:

1) All first-year students will be required to take the WBCT to assess their proficiency levels.
2) English courses will be organized appropriately to students' proficiency.
3) Basic English courses will be designed to help first-year students with limited proficiency.

Essentially, the English instructors should acknowledge the implementation of the WBCT as a placement test so they should help the institution by conducting more research studies to develop sets of WBCT, improving the test to obtain test usefulness or quality of the test, and organizing courses that are appropriate to students' levels of proficiency. According to Bachman and Palmer (1996: 18), test usefulness is a combination of different qualities such as reliability, construct validity, authenticity, interactiveness, impact and practicality.

It is anticipated that the institution inaugurates the WBCT as an English placement test as this test underlying the C-test theory is a proficiency test which assesses integrated skills. In terms of practicality, the Web-based C-Test is more appropriate to be implemented as a placement test, as the test is simply in construction, administration and scoring. Moreover, it is more economical to use the WBCT than other placement tests. It requires less human operation when compared to the Paper and Pen Tests and less computer expertise when compared to Computer-based tests. Finally, test security should be taken into consideration to prevent cheating. That is, several sets of the WBCT should be provided for different administrations and the WBCT should be administered at the institution.

Recommendations for pedagogical implications

The following recommendations are presented to implement the WBCT in the classroom so as to employ the WBCT as a teaching tool to help learners acquire general proficiency of the language.

1. The WBCT can be constructed online and implemented in other types of language assessments; namely, proficiency, diagnostic, and achievement tests.

2. To increase the face validity of the WBCT, it is advisable to construct different texts through the website clozeonline.us to support the learning and teaching in the classroom. This will enhance face validity to Thai learners as well as assist learning and teaching in the classroom. In doing so, students will become familiar with the C-test format and will be able to restore the words without frustration or complaints. Bresnihan & Ray (1994 cited in Sigott, 2004: 56) suggested "informing the test takers in detail about the deletion principle as well as indicating the number of letters deleted by dots or dashes to lessen test anxiety and frustration."

In summary, the WBCT can be used in classrooms as an assessment tool as well as an indirect teaching method. Students will become proficient when they learn the redundancy of the language that will help them understand the whole things. Also, teachers can use the WBCT in different types of language assessment as tests to assess the students' proficiency and language ability.

Recommendations for future research studies

The following recommendations are ideas for future research studies that can be used to enhance the implementations of the WBCT and other modified C-tests in language assessment.

1. Text selection is one of the crucial parts in the construction of the tests. It is advisable that one or two more texts be included in case some texts do not achieve difficulty indices. As such, the texts can be discarded while there are ample texts to continue in the main study.
2. As for the text difficulty, for which this study utilized the Flesch-Kincaid Grade Levels, it corresponds to a U.S. grade level to make it easier to judge the readability level of various books and texts. However, it is obvious that Thai learners' levels are not compatible to the U.S. grade levels. To avoid comparing on Grade Level, it is recommended that the Flesch-Kincaid Reading Ease be used instead. This is because the calculation gives a number indicating that the higher the number, the

easier the text. In addition, it can be simply performed on Microsoft Word. Meanwhile, language teachers or experts can help supervise the appropriateness of the text selection.

3. Likewise, it is recommended that other readability formulas, such as SMOG Index and the Gunning Fog Index, be compiled to observe readability of the texts being used. These formulas can be calculated instantly and are free online. Another example is the Bormuth Grade Level readability formula, which is based on a count of characters rather than syllables per word and words per sentence to determine a grade level.
4. It is advisable to study the test-takers' strategies in taking the test, and an introspective interview or a think-aloud technique is recommended to acquire information more precisely.
5. More samples from different proficiency levels should be employed so that the Web-based C-Test can differentiate the students into more than two levels.
6. To improve other tests for language assessment, the modified C-tests are recommended; for example, the MC-test or the X-Test (the reversed C-test), where the first half of every second word is deleted, and the S-Tests (the Semantic/ Syntactic Tests), where first half or second half of every second content words is deleted.

In summary, the WBCT should be further researched with learners at different ability levels and with texts of varying readability. In the meantime, learning how test-takers employ their knowledge and strategies while taking the tests might give more factual information on how they understand the language. Finally, more research studies on other modified C-tests are advised to extend the use of C-tests as a proficiency test.

REFERENCES

Alderson, J. C. (1979). The effect on the cloze test of changes in deletion frequency. *Journal of Research in Reading*, *2*: 108-118.

Alderson, J. C. (1983). The cloze procedure and proficiency in English as a foreign language. In J. W. Oller, Jr. (ed). *Issues in language testing research*. Rowley: 205-217. Mass.: Newbury House.

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

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: the interface between learning and assessment*. Continuum International Publishing Group.

Babaii, E., & Moghaddam, M. J. (2006). On the interplay between test task difficulty and macro-level processing in the *C-test*. [online] *System, 34* (4): 586-600. Available from: http://www.sciencedirect.com/science [2007, January 20]

Bachman, L. F. (1982). The trait structure of cloze test scores. *TESOL Quarterly, 16 (1)*: 61-70.

Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly, 19 (3)*: 535-56.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Cambridge: Cambridge University Press.

Blum, J. A. (2003). Online C-TESTS for Tailored Training. A pilot post-graduate project proposal, University of Central Florida. [online] Available from: http://clozeonline.us/study.html [2006, July 13]

Boonsathorn, S. (1987). *C-test, proficiency, and reading strategies in ESL*. Unpublished Doctoral Dissertation. University of Alberta, Alberta, Canada.

Boonsathorn, S. (2000). S-Tests: A Rational C-test for Advanced Students. *PASAA: A Journal of Language Teaching and Learning in Thailand,* 40: 65-74.

Boonsathorn, S., Getkham, K., & Boonsathorn, W. (2007). *Development of NIDA Test of English Competence (N-TEC)*. School of Language and Communication, National Institute of Development Administration, Thailand.

Brown, J. D. (1992). Using computers in language testing. *Cross Currents, 19*: 92-99.

Brown, J. D. (1996). *Testing in Language Programs*. Upper Saddle River, New Jersey: Prentice Hall Regents.

Brown, J. D. (2005). *Testing in Language Programs*: *A Comprehensive Guide to English Language Assessment.* New York: McGraw-Hill ESL/ELT.

Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Pearson Education.

Carroll, J. B. (1987). Review of Klein-Braley, C. and Raatz, U. (eds.), C-tests in der Praxis. Fremdsprachen und Hoschscule, AKS-Rundbrief 13/14, Bochum: Arbeitskreis Sprachzentrum [AKS]. *Language Testing, 4*: 99-106.

Cohen, A. C. (1980). *Testing Language Ability in the Classroom*. Rowley, MA: Newbury House.

Cohen, A. (1984). Studying second language learning strategies: how do we get the information? *Applied Linguistics,* 5/2, 131-146.

Cohen, A. D. (1994). *Assessing Language Ability in the Classroom*, 2nd Edition. Boston: Heinie & Heinie Publishers.

Cohen, A.D. (1998). *Strategies in Learning and Using a Second Language.* Harlow: Addison Wesley.

Connelly, M. (1997). Using C-test in English with Postgraduate Students. *English for Specific Purposes, 16*: 139-150.

Daller, H., & Phelan, D. (2006). The C-test and TOEIC as measures of students' progress in intensive short courses in EFL. In R. Grotjahn (ed.) *Der C-test: Theorie, Empirie, Anwendungen/ The C-test: Theory, Empirical Research, Applications. Frankfurt/M.: Lang*.

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing, 9*: 187-206

Dörnyei, Z., & Katona, L. (1993). The C-test: A teacher friendly way to test English language proficiency. [online] *The English Teaching Forum*, *31*: 35-40. Available from: http://exchanges.state.gov/forum/vols/vol31/no1/ p35.htm#page_top [2007, August 22]

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing, 23, (3*): 290-325. SAGE Publications.

Egbert, J. & Hanson-Smith, E. (eds.) (1999). *CALL Environments. Research, Practice and Critical Issues*. Alexandria, VA: TESOL.

Fultz, J. (n.d). Gestalt Psychology. [online] Available from: http://homepages.ius.edu/rallman/gestalt.html [2008, May 12]

Georges, L. (n.d). *C-test 1.2*. [online] Available from: http://l.georges. online.fr/ logiciels.htm. [2006, July 13]

Grotjahn, R. (1987). How to construct and evaluate a C-test: A discussion of some problems and some statistical analyses. In R. Grotjahn, C. Klein-Braley and D. K. Stevenson (eds.): *Taking their measure: The Validity and Validation of Language Tests*: 219-253. Bochum: Brockmeyer.

Grotjahn, R., Klein-Braley C., and Raatz, U. (2002). C-tests: an overview. In J. A. Coleman et al. (eds.). *University Language Testing and the C-test*: 93-114. Bochum: AKS-Verlag.

Grotjahn, R., & Stemmer, B. (1985). C-tests and language processing. In J. A. Coleman et al. (eds.). *University language Testing and the C-test:* 115-130. Bochum: AKS-Verlag, 2002.

Hastings. A. J. (2002a). In Defense of C-testing. In R. Grotjahn (ed.). *Der C-test. Theoretische Grundlagen und praktische Anwendungen*, *4*: 11-25. Bochum, AKS-Verlag.

Hastings. A. J. (2002b). Error analysis of an English C-test: Evidence for integrated processing. In R. Grotjahn (ed.). *Der C-test. Theoretische Grundlagen und praktische Anwendungen*, *4*: 53-66. Bochum, AKS-Verlag.

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

Ikeguchi, C. B. (1998). Do different C-tests discriminate proficiency levels of EL2 learners? [online] *JALT Testing & Evaluation SIG Newsletter, 2*, (1) : 3 - 8. Available from: www.jalt.org/test/ike_2.htm [2005, December 20]

Jafarpur, A. (1995). Is C-testing superior to cloze? *Language Testing*, *12*: 194-216.

Jafarpur, A. (1999). What's Magical about the Rule-of Two for Constructing C-tests? *Language Testing, 30*, (2): 86-100.

Kantahoe, D., Hansombatcharean, R., & Leardvitoonpanya, J. (2003). *The effectiveness of the computer-based semantic/ syntactic test (computer-based s-test).* Unpublished Master's research paper, the National Institute of Development Administration, Bangkok, Thailand.

Kiratibodee, P. (2006). Thai university students' variables as the predictors of computer-based test performance. *PASAA, 38*: 23-43.

Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-test. *Language Testing, 1*: 134-146.

Kozminsky, E., & Kozminsky, L. (2001). How do general knowledge and reading strategies ability relate to reading comprehension of high school students at different educational levels? *Journal of Research in Reading, 24*, *(2)*: 187-204.

Lertudommit, T. (2005). *A Validation of Semantic/ Syntactic tests (S-Tests) for Grade Eight Students at an International School.* Unpublished Master's Degree Research Paper, National Institute of Development Administration, Thailand.

Mathews, T. J. (1988). The Multiple-choice Cloze Test as a Placement Examination: A Critical Analysis. [online] *Delaware Papers in Linguistics*, *1*, *(1)* : 38-53. Available from: http://cc.weber.edu/~tjmathews/home.htm [2006, August 30]

Maulan, S. (2004). Language Testing Using Computers: Examining the Effect of Test-Delivery Medium on Students' Performance. [online] *Internet Journal of e-Language Learning & Teaching*, *1*: 1-14. Available from: http://www.eltrec.ukm.my/ ijellt/pdf/Sumarni.pdf [2005, August 10]

McNamara, T. (2000). *Language Testing.* Oxford: Oxford University Press.

Oller, J.W. (1979). *Language Tests at Schools.* London: Longman.

Poel, C. D., & Weatherly, S. D. (1997). A cloze look at placement testing. [online] *Shiken: JALT Testing & Evaluation SIG Newsletter, 1*, *(1*): 4-10. Available from: http://www.jalt.org/test/po_we.htm [2006, July 13]

Prapphal, K. (1994). A study of the C-test and the X-Test performed by first year science-oriented university students. *PASAA, 24*: 16-23.

Prapphal, K. (1997). Educational Technology for TEFL. *PASAA, 27*: 121-127.

Quick Placement Test CD ROM User Manual. (2001). Oxford University Press.

Quick Placement Test Paper and Pen Test User Manual. (2001). Oxford University Press.

Raatz, U., & Klein-Braley, C. (1981). The C-test–a modification of the cloze procedure. In T. Culhane, C. Klein-Braley, & D.K. Stevenson (eds). *Practice and problems in language testing*: 113-138. Colchester: University of Essex.

Raatz, U., & Klein-Braley, C. (2002). Introduction to language testing and C-tests. In J. A., Coleman, G. Rüdiger & U. Raatz (eds.) *University language Testing and the C-test*: 75-91. Bochum: Brockmeyer. Also at http://www.uni-duisburg.de/FB3/ANGLING/FORSCHUNG/HOWTODO.HTM

Rashid, S. M. (2004). Validating the C-test amongst Malay ESL Learners. [online] Department of English, FBMK, Universiti Putra Malaysia. Available from: http://www.melta.org.my/modules/sections/12.doc [2006, September 20]

Rouhani, M. (2008). Another Look at the C-test: A Validation Study with Iraninan EFL Learners. *The Asian EFL Journal*, *10*, 1: 154-180.

Röver, C. (2001a). Web-based C-tests. In R. Grotjahn (ed). *Der C-test.* Theoretische Grundlagen und praktische Anwendungen, 4: 123-130. Bochum: AKS-Verlag.

Röver, C. (2001b). Web-based Language Testing. [online] *Language Learning & Technology. 5*, 2: 84-94. Available from: http://llt.msu.edu/vol5num2/roever/default.html [2008, April 28]

Rungruangthum, M. (2005). *A Comparative Study between the C-test and the NC-test and between the MC-test and the NMC-test, Using Identical Texts.* Unpublished Master's Thesis, Mahidol University, Thailand.

Saito, Y., Thomas, G. J., & Horwitz, E. K. (1999). Foreign Language Reading Anxiety. *The Modern Language Journal*, *83*, (2): 202-218.

Sasaki, M. (2000). Effects of cultural schemata on students' test taking processes for cloze tests: A multiple data source approach. *Language Testing*, *17*: 85-114.

Sigott, G. (2002). High-level processes in C-test taking? In R. Grotjahn (ed.). *Der C-test. Theoretische Grundlagen und praktische Anwendungen, 4*: 67-82. Bochum: AKS-Verlag.

Sigott, G. (2004). *Towards Identifying the C-test Construct.* Frankfurt: Peter Lang.

Stansfield, C. (1980). The Cloze Procedure as a Progress Test. [online] *Hispania*, *63*, 4: 715-718. Available from: http://www.jstor.org/pss/340752 [2007, July 13]

Stevenson, J., & Gross, S. (1991). Use of a Computerized Adaptive Testing Model for ESOL/bilingual entry/exit decision making. In P. Dunkel (ed.). *Computer-assisted language learning and testing: Research issues and practice*: 223-235. New York: Newbury House.

Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly, 30*: 415-433.

Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). The relationship between computer familiarity and performance on computer-based TOEFL test tasks. Research Report 61, Princeton. NJ: Educational Testing Service. In R. Grotjahn (ed) 2002. *Der C-test. Theoretische Grundlagen und praktische Anwendungen, 4*: 123-130. Bochum: AKS-Verlag.

UCLA Dept. of Applied Linguistics/TESL and Center for Digital Humanities, (2001-2003) Limitations of Paper & Pencil Tests.[online] Available from: http://www.weblas.ucla.edu/ [2006, July 13]

Yamashita, J. (2003). Process of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, 20, 3: 267-293.

APPENDICES

Appendix A

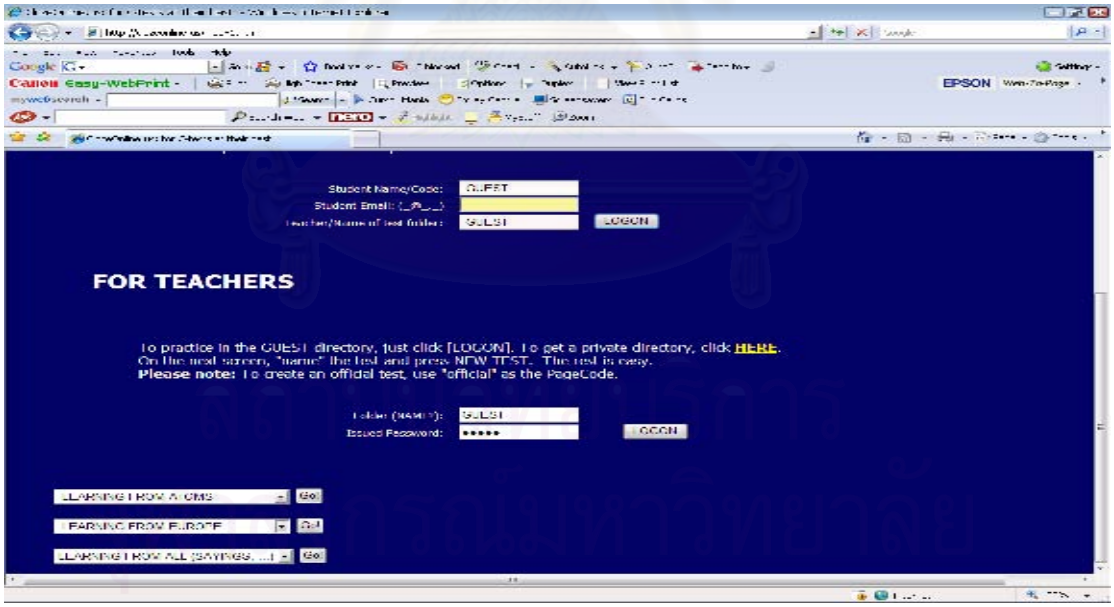
Construction Process of the Web-based C-Test (Adapted from www.clozeonline.us)

In constructing the Web-based C-Test, there are two different processes: the process of making the tests for the instructors, and the process of doing the tests for the students or test-takers.

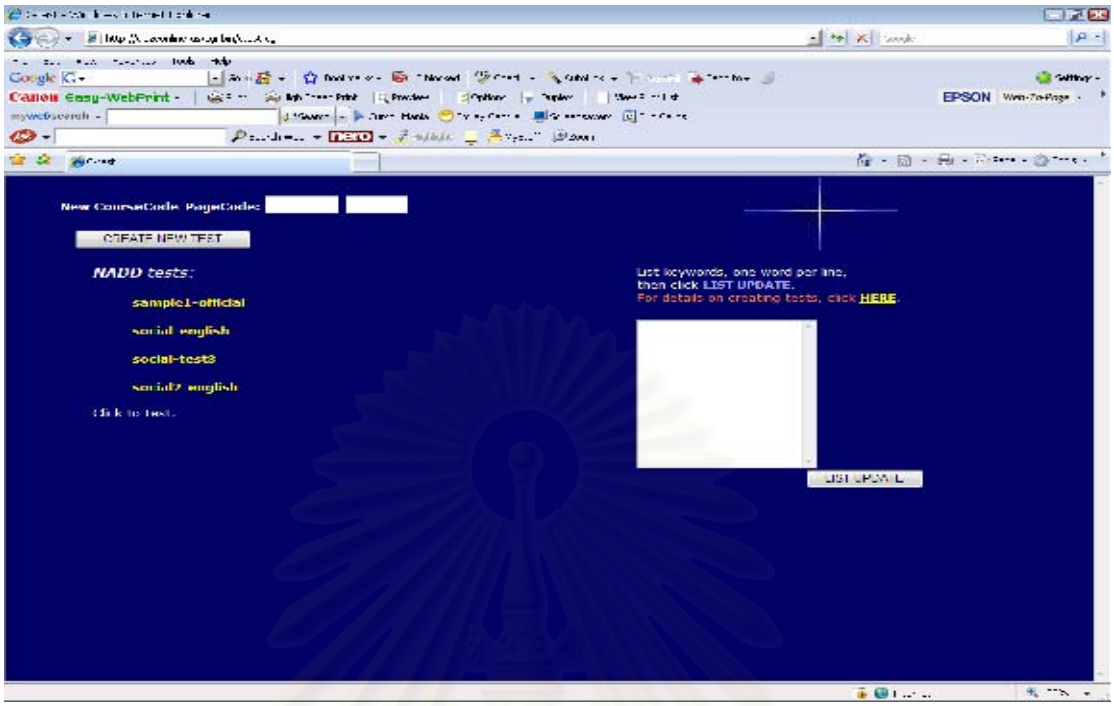
The process of making the WBCT for instructors

The process of making the tests was adapted from the http://clozeonline.us site, the program used for this study. There are seven steps in constructing the WBCT as follows:

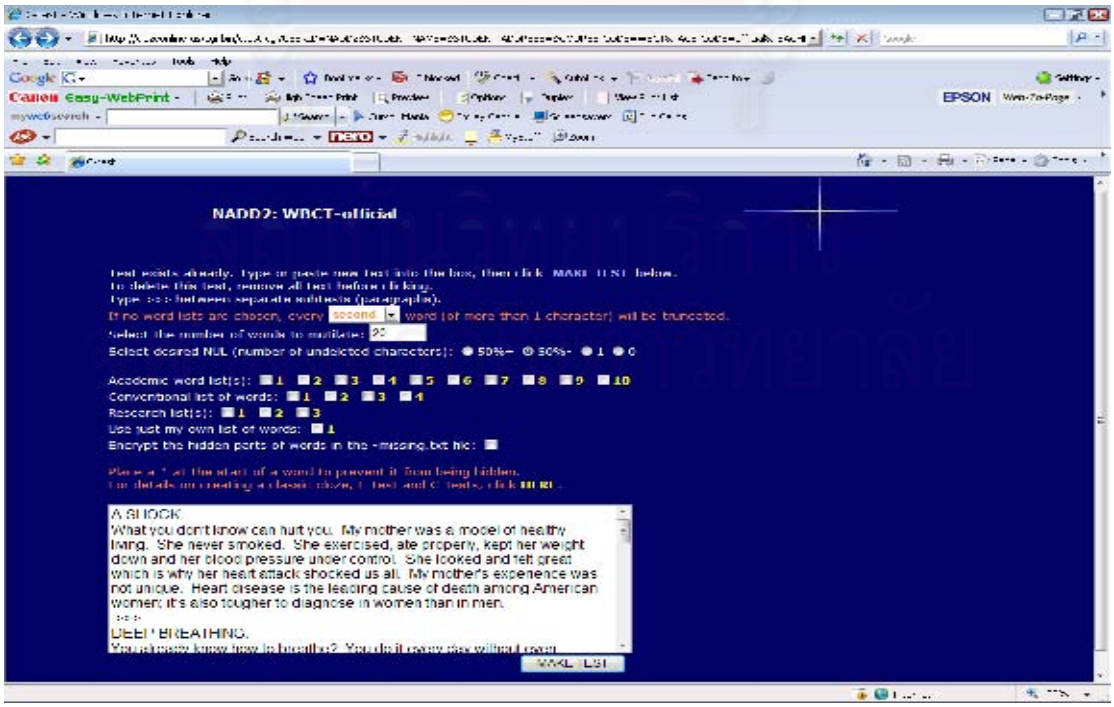
1. Go to "http://clozeonline.us for teachers" and type in the user name and password, which are supplied by the webmaster, then press 'LOG ON.'



2. At the "Create New Test" screen, name the test with two strings; e.g., a course and page code, then click on "Create New Test."



3. At the “Make Test” screen, the name of the test appears on the screen. Then, copy the selected passages from a file and paste them into the box, with this string >>> between each passage so that the passage will appear to the students randomly one at a time.



4. Change the defaults to a variety of combinations or leave the settings at second word deletion with the number of word deletions set to 20 and click 'MAKE TEST.'
5. In details, at the http://clozeonline.us. for teachers, the instructions can be read as follows:

Example 2.2

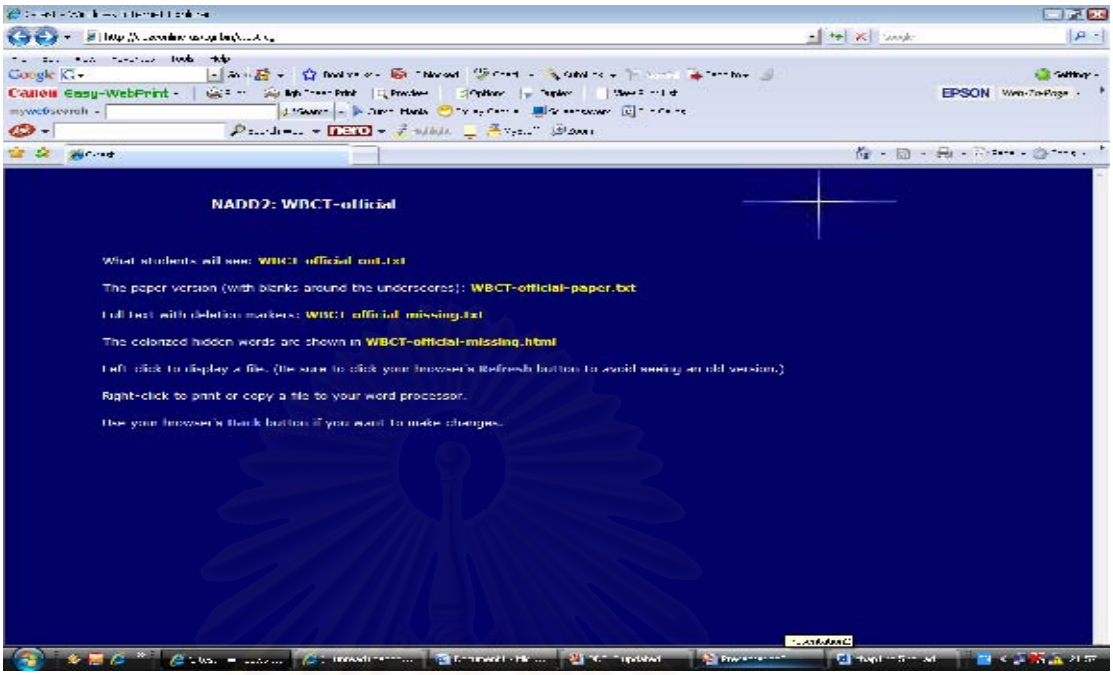
Instructions for making the WBCT

- New Test. Type or paste your text into the box, then click **MAKE TEST** below.
  Type >>> between separate subtests (paragraphs).

- If no word lists are chosen, every second word (of more than 1 character) will be truncated.
  Select the number of words to
  mutilate:
  Select NUL (number of undeleted characters): 50%+ 50%- 1 0
- Academic word list(s): **1 2 3 4 5 6 7 8 9 10**
- Conventional list of words: **1 2 3 4**
- Research list(s): **1 2 3**
- Use just my own list of words: **1**
- Encrypt the hidden parts of words in the -missing.txt file:

Place an asterisk * at the start of a word to prevent it from being hidden.
For details on creating a classic cloze, L-test and C-Tests, click **HERE**.

6. After clicking on the 'MAKE TEST,' the test appears in 4 different versions: a) what the students will see, b) the paper version with blanks around the underscores, c) full text with deletion markers, and d) the colorized hidden words.



The colorized or bold-typed hidden words help the researcher to see the truncated words clearly. It is, therefore, very useful for the researcher to check the words. The bold-typed hidden words are displayed as follows:

A SHOCK.

What you don't know can hurt you. My mot**her** was a mo**del** of hea**lthy** living. S**he** never smo**ked**. She exer**cised**, ate prop**erly**, kept h**er** weight do**wn** and h**er** blood pres**sure** under con**trol**. She loo**ked** and fe**lt** great wh**ich** is w**hy** her he**art** attack sho**cked** us a**ll**. My moth**er's** experience was not unique. Heart disease is the leading cause of death among American women; it's also tougher to diagnose in women than in men.

>>>

DEEP BREATHING.

You already know how to breathe? You d**o** it ev**ery** day wit**hout** even thin**king** about i**t**. Your brea**thing** technique i**s** not a**s** healthy a**s** you mi**ght** think. Mo**st** of u**s** breathe t**oo** shallowly a**nd** too qui**ckly**. Our lu**ngs** and he**art** would gre**atly** prefer lon**ger**, slower, dee**per** breaths. This is true for general health, and it is true for managing stress. Deep breathing helps drive away the reaction when we're stressed. It sends a

signal to your brain to slow down, which results in hormonal and physiological changes that slow heart rate and lower blood pressure.
>>>
LAUGHTER.
What is the greatest reward of being alive? Is it chocolate, ice cream, vacations, a perfect night's sleep, or the satisfaction of a job well done? Ask a thousand people and you'll get as many answers. But one true pleasure that covers all people is laughter. No matter your age, wealth, race, or living situation, life is good when laughter is frequent. Life is also healthier. Research finds that humor can help you cope better with pain, enhance your immune system, reduce stress, even help you live longer.
>>>

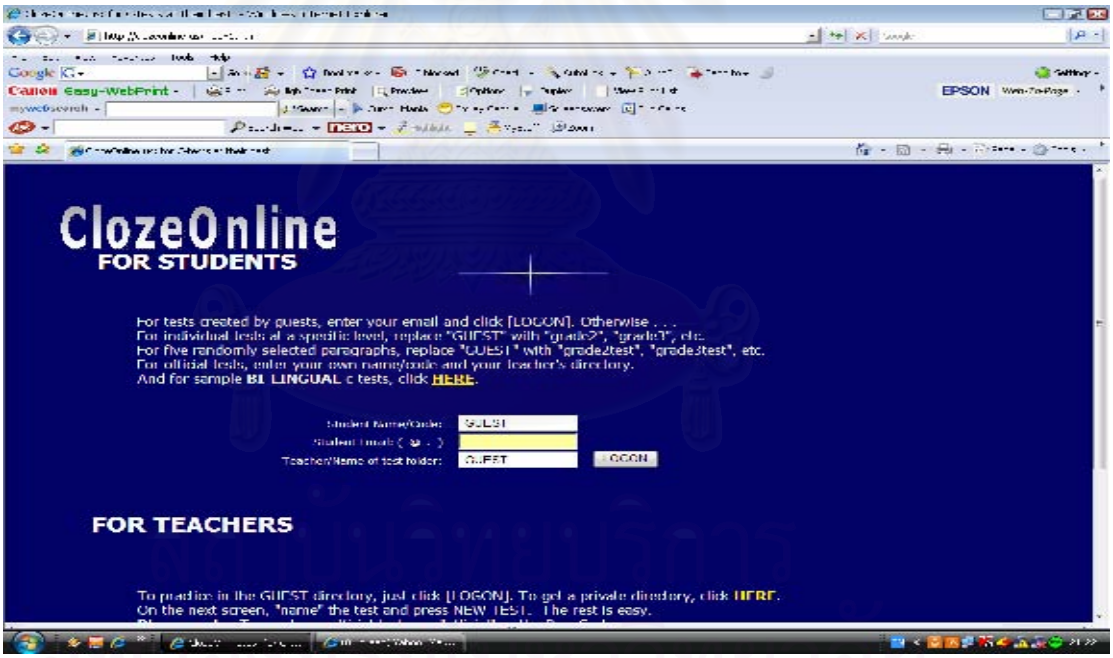
INTELLIGENCE PILLS.
Some scientists have predicted that healthy adults and children may one day take drugs to improve their intelligence. A research group has suggested that such drugs might become as common as coffee or tea within the next couple of decades. In the future, students taking exams might have to take drugs tests like athletes. There are already drugs that are known to improve mental performance, like Ritalin, which is given to children with problems concentrating. A drug given to people who have trouble sleeping also helps people remember numbers.
>>>
MOVIE REVIEWS.
Movie reviews are written, not to tell you what to see, but to help you decide whether or not you would like to see a certain film. Movie critics offer their opinions on the qualities of a particular movie. Sometimes their reviews are positive and other times they are negative. Different reviewers may have very different opinions about the same movie. What one movie reviewer finds entertaining, another may find disappointing.

7. In the meantime, the newly constructed C-Tests are automatically listed on the http://clozeonline.us for students' screen. The test is now ready for students to log on to do the tests.

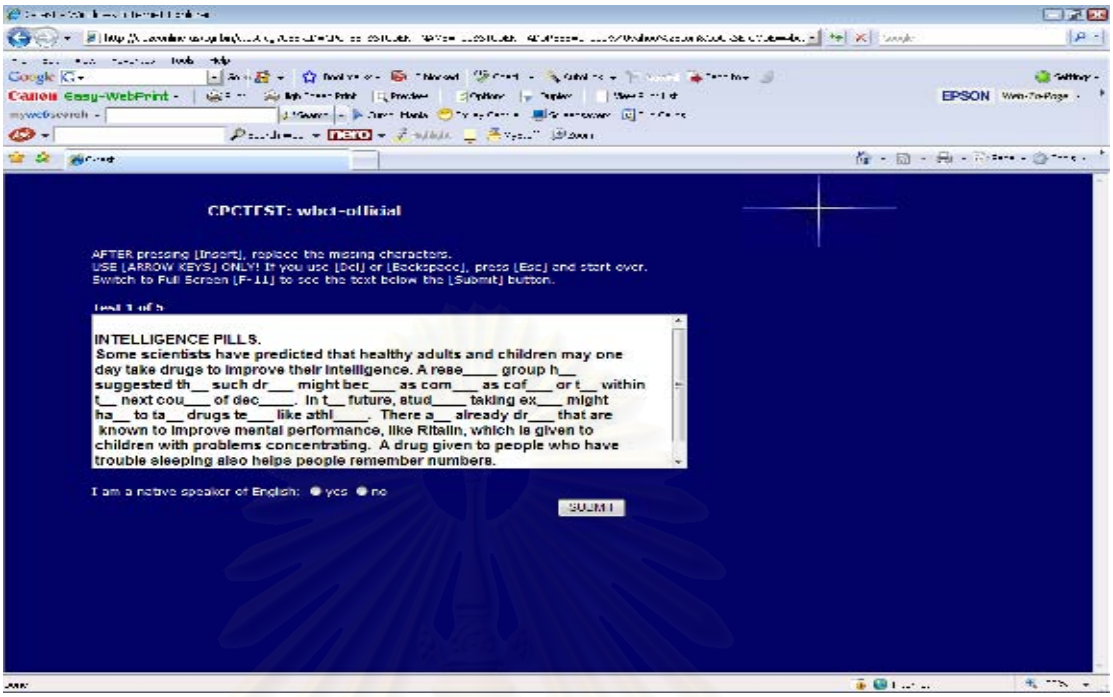
The process of doing the WBCT for test-takers

The process of doing the test for test-takers was adapted from the software program to be implemented in this study. The test-takers are required to follow the following steps:

1. The student logs on to "http://clozeonline.us for students" and enters his/her name at [Student Name] followed by an email address in the format of "name@name.com."



2. This is followed by entering the [Teacher's Name], in my case, "NADD" for the sample test and "CPCTEST" for the actual test, WBCT," and by clicking [LOGON].
3. When the student clicks on the name of the test, in my case, "SAMPLE-official" for the tutoring period or "WBCT-official" for the actual test, the test appears one paragraph at a time.



4. Before doing the test, the student needs to follow the instructions on doing the WBCT. The researcher stresses that the student presses the keyboard's [Insert] key and activate (click) the curser inside the text box before using the arrow keys to maneuver the curser to the beginning of the first underscore. The instructions can be read as follows:

Example 2.3

Instructions for doing the WBCT

CPCTEST: WBCT-official

AFTER pressing [Insert], replace the missing characters.
USE [ARROW KEYS] ONLY! If you use [Del] or [Backspace], press [Esc] and start over.
Switch to Full Screen [F-11] to see the text below the [Submit] button.

5. For the sample test, the researcher acts as a coach to familiarize the student with the test format and the process in doing the WBCT.
6. The student clicks on 'no' to the statement under the test that says: *I am a native speaker of English: yes no.*
7. The student is given 9-10 minutes to restore each text before clicking on the SUBMIT button. The results after submission of the test appear at the bottom as displayed:

Example 2.4

The results on screen after submission

> Your score is currently 8 out of 20. Replace more underscores and **SUBMIT** again,
> or **skip ahead to the next test in this series,**
> **or end the tests now. (*You must click to get credit.*)**
> **Thank you for aiding our research.**

8. For this study, students get only one try, therefore, the student clicks on *'Skip ahead to the next test in this series'* at the bottom of the screen. Another test appears. This process repeats until the student finishes the five subtests provided.
9. As the student submits each test, the results are displayed on the screen as well as sent to the researcher's e-mail.
10. When working with this online version of the C-test, it is important to stress the pressing of the [Insert] key after every time the SUBMIT button is pressed. Should they forget (which is evidenced by seeing text being pushed out of the way), they could use the [Esc] key to restore the window to its proper integrity.
11. The time allotment for the WBCT is 45-50 minutes for restoring the five subtests.

Appendix B

Questionnaire for three raters

Dear …….

I am an EIL student, Batch 4, conducting a study on the Web-based C-Test aiming to implement it as an English placement test for Rajamangala University of Technology, Tawan-ok, Chakrabongse Bhuvanath campus. The title is *"The Use of the Web-based C-Test (WBCT) as an English Placement Test for Rajamangala University of Technology Tawan-ok and an Investigation of Test-takers' Strategies."*

The purpose of this questionnaire is to obtain your opinions towards the 10 prospective passages which will be implemented in the Web-based C-Test in terms of these characteristics: general, authentic, accurate, appropriate and interesting so as to avoid biases on passage selection. Included are sources, word counts and ascending readability grade levels, ranging from 5.4 – 11.8 using Flesch-Kincaid Grade Level.

After your rating, five short passages out of ten will be selected for the construction process of the Web-based C-Test to be implemented in this study. In brief, the Web-based C-Test will consist of five subtests and each passage will have 20 deletions. The deletions start from the second sentence where half of every second word is deleted. The deletion will total 100 items. The construction and the administration of the Web-based C-Test will be performed online at http://clozeonline.us.

The subjects of the study are first-year students at Rajamangala University of Technology Tawan-ok, at Chakrabongse Bhuvanath Campus in Bangkok in the academic year of 2007.

Please find the questionnaire to rate on the 10 passages. The 10 short passages in full texts and in C-Test format are attached in appendix. Thank you very much for your cooperation.

Naddhaporn Kammasorn
EIL student, Chalalongkorn University

QUESTIONNAIRE

Please check [√] in the blanks according to your opinions to each passage.

Passage	Word counts	Read-ability	Sources	Characteristics of the passages	Yes	No
1. A SHOCK	73	5.4	"Editor's Note". *National Geographic*. Feb 2007, Vol 211, No. 2: 6	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……
2. DEEP BREATH-ING	95	6.4	Health, S. "The Right Breath", *Reader's Digest*. http://www.rd.com/content/openContent.do?contentId=16354	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……
3. AUTUMN	95	6.7	http://www.teach-nology.com/worksheets/language_arts/reading_comp/elem/ver12/	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……
4. LAUGHTER	88	7.0	"Enhance Your Sense of Humor." *Reader's Digest*. http://www.rd.com/content/openContent.do?contentId=16125	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……
5. TAKING PICTURES	62	7.3	Icenhower, M. W. "Taking a Picture." *Ambassador Youth*. http://www.thercg.org/youth/articles/0403-tap.html	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……
6. FISH AND WHALES	92	8.4	Gander, J. C. 2003. Fish and Whales. http://www.stemnet.nf.ca/CITE/whale_fish.pdf	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……
7. INTELLI-GENCE PILLS	88	9.9	English Language Articles. 2007.*Intelligent pills*. http://www.usingenglish.com/comprehension/21.html	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……
8. BARBECUE	78	10.3	Smith, A. F. editor. *2004. "Barbecue."* Oxford Encyclopedia of Food and Drink in America. Oxford University Press:New York. Volume 1: 64-65.	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……
9. MOVIE REVIEWS	71	11.3	Reading Movie Reviews." *Bangkok Post*. http://bangkokpost.net/education/movies.htm	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……

Passage	Word counts	Read-ability	Sources	Characteristics of the passages	Yes	No
10. STUDY IN THE U.S	79	11.8	Barrett, M. "Is the USA Right for You?" *TOEFL access: US edition*. http://www.toeflaccess.com/articles/ETS/archive/us/	1. general 2. authentic 3. accurate 4. appropriate 5. interesting	…… …… …… …… ……	…… …… …… …… ……

Comments and suggestions:
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………… …….
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………..
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………..
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………..
………………………………………………………………………………………….

Appendix (for the questionnaire)

A. The 10 short passages in full texts

1. A SHOCK

What you don't know can hurt you. My mother was a model of healthy living. She never smoked. She exercised, ate properly, kept her weight down and her blood pressure under control. She looked and felt great – which is why her heart attacked shocked us all. My mother's experience was not unique. Heart disease is the leading cause of death among American women; it's also tougher to diagnose in women than in men.

(73 words, Readability 5.4)

Source: "Editor's Note". *National Geographic*. Feb 2007, Vol 211, No. 2 page 6

>>>

2. DEEP BREATHING

You already know how to breathe? You do it every day without even thinking about it. Your breathing technique is not as healthy as you might think. Most of us breathe too shallowly and too quickly. Our lungs and heart would greatly prefer longer, slower, deeper breaths. This is true for general health, and it is true for managing stress. Deep breathing helps drive away the reaction when we're stressed. It sends a signal to your brain to slow down, which results in hormonal and physiological changes that slow heart rate and lower blood pressure.

(95 words, Readability 6.4)

Source: Health, S. "The Right Breath", *Reader's Digest*. Retrieved July 13, 2006 from http://www.rd.com/content/openContent.do?contentId=16354

>>>

3. AUTUMN.

Autumn is the season that falls between summer and winter. There are many changes that begin in this fascinating season. Days become shorter. Leaves of trees turn from green to red, yellow and orange. Trees need sunlight to keep their leaves a lively green. Without sunlight leaves turn colors. The grass is blanketed with frost, almost every morning, as temperatures reach the freezing point. Animals start storing up a food

supply to last the long winter months. These changes occur as we adjust from the heat of the summer to the chill of the winter.

(95 words, Readability 6.7)

Source: Autumn. (n.d) Retrieved August 13, 2006 from http://www.teach-nology.com/worksheets/language_arts/reading_comp/elem/ver12/

>>>

4. LAUGHTER

What is the greatest reward of being alive? Is it chocolate, ice cream, vacations, a perfect night's sleep, or the satisfaction of a job well done? Ask a thousand people and you'll get as many answers. But one true pleasure that covers all people is laughter. No matter your age, wealth, race, or living situation, life is good when laughter is frequent. Life is also healthier. Research finds that humor can help you cope better with pain, enhance your immune system, reduce stress, even help you live longer.

(88 words, Readability 7.0)

Source: "Enhance Your Sense of Humor." *Reader's Digest*. Retrieved July 13, 2006 from http://www.rd.com/content/openContent.do?contentId=16125

>>>

5. TAKING PICTURES

Do you ever think back to occasions when you shared fun times with others? As you reflect on people and places, do you sometimes find it difficult to recall the details? Do you try to remember faces, but too often only have a foggy picture in your mind? Taking snapshots can fill in the details and make past events seem real again!

(62 words, Readability 7.3)

Source: Icenhower, M. W. "Taking a Picture." *Ambassador Youth*. Retrieved Nov 22, 2006 from http://www.thercg.org/youth/articles/0403-tap.html

>>>

6. FISH AND WHALES

Fish and whales are both vertebrates. This means they both have backbones. They also live in aquatic environments. Except for a couple of species, whales live only in the ocean water. Fish, however, inhabit both fresh and salt waters. Whales are among the largest animals on earth. Fish are among the smallest. Being mammals, whales are warm-blooded. This means they must maintain a constant and warm body temperature.

Fish, on the other hand, are more like reptiles. Most are cold-blooded. This means they change their body temperature to match the surrounding water.

(92 words, Readability 8.4)

Source: Gander, J. C. 2003. Fish and Whales. Retrieved September 1, 2007 from http://www.stemnet.nf.ca/CITE/whale_fish.pdf

>>>

7. INTELLIGENCE PILLS

Some scientists have predicted that healthy adults and children may one day take drugs to improve their intelligence. A research group has suggested that such drugs might become as common as coffee or tea within the next couple of decades. In the future, students taking exams might have to take drugs tests like athletes. There are already drugs that are known to improve mental performance, like Ritalin, which is given to children with problems concentrating. A drug given to people who have trouble sleeping also helps people remember numbers.

(89 words, Readability 9.9)

English Language Articles. 2007. *Intelligent pills*. Retrieved September 1, 2007 from http://www.usingenglish.com/comprehension/21.html

>>>

8. BARBECUE.

Barbecue is a method of slow-cooking meat over coals, also known a barbeque, bar-b-q, BBQ, or simply cue. Europeans had been cooking meat over fires for thousands of years. It was the low heat of the coals and the slowness of the process that set the New World method apart. The Europeans quickly adopted this method of slow cooking, discovering fairly early that hogs made great barbecue. Barbecue parties featuring whole hogs became fashionable by the late 1600s.

(78 words, Readability 10.3)

Source: Smith, A. F. editor. *2004. "Barbecue."* Oxford Encyclopedia of Food and Drink in America. Oxford University Press:New York. Volume 1: 64-65. Retrieved August 8, 2007 from http://www.foodtimeline.org/foodfaq7.html#barbeque

>>>

9. MOVIE REVIEWS

Movie reviews are written, not to tell you what to see, but to help you decide whether or not you would like to see a certain film. Movie critics offer their opinions on the qualities of a particular movie. Sometimes their reviews are positive and other times they are negative. Different reviewers may have very different opinions about the same movie. What one movie reviewer finds entertaining, another may find disappointing.

(71 words, Readability 11.3)

Source: "Reading Movie Reviews." *Bangkok Post*. Retrieved June 13, 2006 from http://bangkokpost.net/education/movies.htm

>>>

10. STUDY IN THE U.S

An international student has plenty to consider before pursuing a U.S. academic degree. According to the *EducationUSA Web site, international students trying to decide whether to study in the United States need to define goals for their job first. This will help the student find the right program at the right school. It's important to decide on a program of study so the international student can make sure the U.S. degree will be accepted in his or her country.

(79 words, Readability 11.8)

Barrett, M. "Is the USA Right for You?" *TOEFL access: US edition*. Retrieved Jan 1, 2007 form http://www.toeflaccess.com/articles/ETS/archive/us/study/is_usa_right.html

>>>

B. The 10 short passages in C-Test format

1. A SHOCK
What you don't know can hurt you. My mot___ was a mo___ of hea____ living. S__ never smo___. She exer_____, ate prop____, kept h__ weight do__ and h__ blood pres____ under con____. She loo___ and fe__ great -- wh___ is w__ her he___ attacked sho____ us a__. My moth____ experience was not unique. Heart disease is the leading cause of death among American women; it's also tougher to diagnose in women than in men.
>>>

2. DEEP BREATHING
You already know how to breathe? You d_ it ev___ day wit____ even thin____ about i_. Your brea_____ technique i_ not a_ healthy a_ you mi___ think. Mo__ of u_ breathe t__ shallowly a__ too qui____. Our lu___ and he___ would gre____ prefer lon___, slower, dee___ breaths. This is true for general health, and it is true for managing stress. Deep breathing helps drive away the reaction when we're stressed. It sends a signal to your brain to slow down, which results in hormonal and physiological changes that slow heart rate and lower blood pressure.
>>>

3. AUTUMN

Autumn is the season that falls between summer and winter. There a__ many cha____ that be___ in th__ fascinating sea___. Days bec___ shorter. Lea___ of tr___ turn fr__ green t_ red, yel___ and ora___. Trees ne__ sunlight t_ keep th___ leaves a liv___ green. Wit____ sunlight lea___ turn col___. The gr___ is blanketed with frost, almost every morning, as temperatures reach the freezing point. Animals start storing up a food supply to last the long winter months. These changes occur as we adjust from the heat of the summer to the chill of the winter.
>>>

4. LAUGHTER
What is the greatest reward of being alive? Is i_ chocolate, i__ cream, vaca_____, a per____ night's sl___, or t__ satisfaction o_ a j__ well do__? Ask a thou____ people a__ you'll g__ as ma__ answers. B__ one tr__ pleasure th__ covers a__ people i_ laughter. N_ matter yo__ age, wealth, race, or living situation, life is good when laughter is frequent. Life is also healthier. Research finds that humor can help you cope better with pain, enhance your immune system, reduce stress, even help you live longer.
>>>

5. TAKING PICTURES

Do you ever think back to occasions when you shared fun times with others? As y__ reflect o_ people a__ places, d_ you some_____ find i_ difficult t_ recall t__ details? D_ you t__ to reme____ faces, b__ too of___ only ha__ a fo___ picture i_ your mi__? Taking snap_____ can fi__ in t__ details and make past events seem real again!
>>>

6. FISH AND WHALES

Fish and whales are both vertebrates. This me___ they bo__ have back_____. They al__ live i_ aquatic enviro_____. Except f__ a cou___ of spe____, whales li__ only i_ the oc___ water. Fi__, however, inh____ both fr___ and sa__ waters. Wha___ are am___ the lar____ animals o_ earth. Fish are among the smallest. Being mammals, whales are warm-blooded. This means they must maintain a constant and warm body temperature. Fish, on the other hand, are more like reptiles. Most are cold-blooded. This means they change their body temperature to match the surrounding water.
>>>

7. INTELLIGENCE PILLS
Some scientists have predicted that healthy adults and children may one day take drugs to improve their intelligence. A rese____ group h__ suggested th__ such dr___ might bec___ as com___ as cof___ or t__ within t__ next cou___ of dec____. In t__ future, stud____ taking ex___ might ha__ to ta__ drugs te___ like athl____. There a__ already dr___ that are known to improve mental performance, like Ritalin, which is given to children with problems concentrating. A drug given to people who have trouble sleeping also helps people remember numbers.
>>>

8. BARBECUE

Barbecue is a method of slow-cooking meat over coals, also known a barbeque, bar-b-q, BBQ, or simply cue. Europeans h__ been coo____ meat ov__ fires f__ thousands o_ years. I_ was t__ low he__ of t__ coals a__ the slow____ of t__ process th__ set t__ New Wo___ method ap___. The Euro_____ quickly ado____ this met___ of sl__ cooking, discovering fairly early that hogs made great barbecue. Barbecue parties featuring whole hogs became fashionable by the late 1600s.
>>>

9. MOVIE REVIEWS

Movie reviews are written, not to tell you what to see, but to help you decide whether or not you would like to see a certain film. Movie cri____ offer th___ opinions o_ the qual_____ of a parti_____ movie. Some_____

their rev____ are posi____ and ot___ times th__ are nega____. Different revi_____ may ha__ very diff_____ opinions ab___ the sa__ movie. Wh__ one mo___ reviewer fi___ entertaining, ano____ may find disappointing.
>>>

10. STUDY IN THE US

An international student has plenty to consider before pursuing a US academic degree. According t_ the Educat_____ Web si__, international stud____ trying t_ decide whe____ to st___ in t__ United Sta___ need t_ define go___ for th___ job fi___. This wi__ help t__ student fi__ the ri___ program a_ the ri___ school. It__ important to decide on a program of study so the international student can make sure the U.S. degree will be accepted in his or her country.

Appendix C

Retrospective Interview Form

Retrospective Interview Form (for researcher only)		
Date…….	**Time……..**	
Researcher: …….	**Level: 0 1 2 3 4 5**	
A. Test-Taking Strategies		
Questions	**Answers**	**Remarks**
I. What were your strategies in doing the test?		
1. Reading the clause		
2. Reading the sentence		
3. Scanning the whole text		
4. Reading each word or parts of sentences		
5. Using extra knowledge		
6. Using background knowledge		
7. Guessing		
8. Finding answer in the passage		
9. Analyzing the passage		
10. Translating words or sentences in the context		
11. Using grammar knowledge		
12. Others ……..... ……...		
II. What in the passage helped you restore the words most?		
6. The title of the passage		
2. The first sentence (left standing)		
3. The first half of the word.		
4. The rest of the text		
5. The word restored		

Retrospective Interview Form (for researcher only) Page 2		
Date…….	**Time……..**	
Researcher: …….	**Level: 0 1 2 3 4 5**	
A. Test-Taking Strategies		
Questions	**Answers**	**Remarks**
6. Others …… ……..		
III. What kinds of words were most difficult to restore?		
1. 2-3 letters (short words)		
2. 4-6 letters. (medium length words)		
3. More than 6 letters (long words)		
4. Others ……. ……		
IV. What knowledge most helped you restore the words?		
1. Vocabulary.		
2.Grammar		
3.Reading comprehension		
4.General proficiency		
5.Others …… ……..		
V. Did you understand all the five subtests after restoring the words?		
1. Yes, most of them. (3-5).		
2. Yes, a few. (1-2)		
3. Not at all.		
4. Others…… ……		

Retrospective Interview Form (for researcher only) Page 3		
B. Opinions		
Questions	**Answers**	**Remarks**
I. What do you think the test tests?		
1. General English ability		
2. Vocabulary		
3. Reading comprehension		
4. Grammar		
5. Others …… ……		
II. Do you like doing the test on the web-based delivery?		
1. Yes.		
2. No.		
3. Others …… ……		
III. Do you think the test can assess your English proficiency?		
1. Yes.		
2. No.		
3. Others …… ……		
IV. What is your opinion towards the test?		
1. Challenging.		
2. Difficult.		
3. Easy.		
4. Boring.		
5. Others …….. ……		

Questions	Answers	Remarks
V. What do you think about the time allotment?		
1. Sufficient.		
2. Too short.		
3. Too long.		
4. Others …….. ……..		

Appendix D

Passages in full texts and in C-Test format in Pilot Study 1

Passage 1 in Pilot Study 1

Passage in full text	Passage in C-Test format
1. TIME IS PRECIOUS The attacks on September 11 have given meaning to people not only in the United States but also around the world. I have learned that life is precious and at any given moment, you could lose the love of your life, a best friend, a parent, or anyone else; and you have to live each and every day as if it could be your last. Live it up and be happy. Smile and laugh at least once a day. And make someone else smile and laugh. They say when you're about to die, your whole life flashes before your eyes. (100 words, Readability = 5.6) Source: Gordon, G. M. "Teen Give Thanks" *Reader's Digest*. Retrieved from http://www.rd.com/content/openContent.do?contentId=15304	1. TIME IS PRECIOUS The attacks on September 11 have given meaning to people not only in the United States but also around the world. I ha__ learned th__ life i_ precious a__ at a__ given mom___, you co___ lose t__ love o_ your li__, a be__ friend, a par___, or any___ else; a__ you ha__ to li__ each a__ every d__ as i_ it co___ be your last. Live it up and be happy. Smile and laugh at least once a day. And make someone else smile and laugh. They say when you're about to die, your whole life flashes before your eyes.

Passage 2 in Pilot Study 1

Passage in full text	Passage in C-Test format
2. DEEP BREATHING You already know how to breathe? You do it every day, without even thinking about it. Your breathing technique is not as healthy as you might think. Most of us breathe too shallowly, and too quickly. Our lungs and heart would greatly prefer longer, slower, deeper breaths. This is true for general health, and it is true for managing stress. Deep breathing helps drive away the reaction when we're stressed. It sends a signal to your brain to slow down, which results in hormonal and physiological changes that slow heart rate and lower blood pressure. (95 words, Readability = 6.4) Source: Health, S. "The Right Breath", *Reader's Digest*. Retrieved from http://www.rd.com/content/openContent.do?contentId=16354	2. DEEP BREATHING You already know how to breathe? You d_ it ev___ day, wit____ even thin____ about i_. Your brea_____ technique i_ not a_ healthy a_ you mi___ think. Mo__ of u_ breathe t__ shallowly, a__ too qui____. Our lu___ and he___ would gre____ prefer lon___, slower, dee___ breaths. This is true for general health, and it is true for managing stress. Deep breathing helps drive away the reaction when we're stressed. It sends a signal to your brain to slow down, which results in hormonal and physiological changes that slow heart rate and lower blood pressure.

Passage 3 in Pilot Study 1

Passage in full text	Passage in C-Test format
3. LAUGHTER What is the greatest reward of being alive? Is it chocolate, ice cream, vacations, a perfect night's sleep, or the satisfaction of a job well done? Ask a thousand people and you'll get as many answers. But one true pleasure that covers all people is laughter. No matter your age, wealth, race, or living situation, life is good when laughter is frequent. Life is also healthier. Research finds that humor can help you cope better with pain, enhance your immune system, reduce stress, even help you live longer. (88 words, Readability = 7.0) Source: "Enhance Your Sense of Humor." *Reader's Digest*. Retrieved from http://www.rd.com/content/openContent.do?contentId=16125	3. LAUGHTER What is the greatest reward of being alive? Is i_ chocolate, i__ cream, vaca_____, a per____ night's sl___, or t__ satisfaction o_ a j__ well do__? Ask a thou____ people a__ you'll g__ as ma__ answers. B__ one tr__ pleasure th__ covers a__ people i_ laughter. N_ matter yo__ age, wealth, race, or living situation, life is good when laughter is frequent. Life is also healthier. Research finds that humor can help you cope better with pain, enhance your immune system, reduce stress, even help you live longer.

Passage 4 in Pilot Study 1

Passage in full text	Passage in C-Test format
4. TREES Almost everyone loves trees, but did you know that trees can help people heal faster? Hospital patients have been shown to recover faster when offered a view with trees. And while trees certainly benefit any landscape, they are actually important to our daily well-being. Trees improve the quality of the air around us by absorbing dust. Their leaves also absorb carbon dioxide, ozone, carbon monoxide, and sulfur dioxide, before we ever inhale them. Plus, trees protect us from the effects of the sun and from rain, hail, and snow and release clean oxygen for us to breathe. (97 words, Readability = 9.4) Source: Zimmerman, S. "Plant a Tree in Your Community." *Reader's Digest.* Retrieved from http://www.rd.com/content/openContent.do?contentId=29483	4. TREES Almost everyone loves trees, but did you know that trees can help people heal faster? Hospital pati____ have be__ shown t_ recover fas___ when off____ a vi__ with tr___. And wh___ trees cert_____ benefit a__ landscape, th__ are actu____ important t_ our da___ well-being. Tr___ improve t__ quality o_ the a__ around u_ by abso_____ dust. Their leaves also absorb carbon dioxide, ozone, carbon monoxide, and sulfur dioxide, before we ever inhale them. Plus, trees protect us from the effects of the sun and from rain, hail, and snow and release clean oxygen for us to breathe.

Passage 5 in Pilot Study 1

Passage in full text	Passage in C-Test format
5. MOVIE REVIEWS Movie reviews are written, not to tell you what to see, but to help you decide whether or not you would like to see a certain film. Movie critics offer their opinions on the qualities of a particular movie. Sometimes their reviews are positive and other times they are negative. Different reviewers may have very different opinions about the same movie. What one movie reviewer finds entertaining, another may find disappointing. (71 words, Readability = 11.3) Source: "Reading Movie Reviews." *Bangkok Post*. Retrieved from http://bangkokpost.net/education/movies.htm	5. MOVIE REVIEWS Movie reviews are written, not to tell you what to see, but to help you decide whether or not you would like to see a certain film. Movie cri____ offer th___ opinions o_ the qual_____ of a parti_____ movie. Some_____ their rev____ are posi____ and ot___ times th__ are nega____. Different revi_____ may ha__ very diff_____ opinions ab___ the sa__ movie. Wh__ one mo___ reviewer fi___ entertaining, ano____ may find disappointing.

Appendix E

International Congruence Agreement (ICA) of passage selection and comments from three raters

Table 1

International Congruence Agreement (ICA) of passage selection and comments from three raters

Passage	Characteristics of the passages	Rater 1	Rater 2	Rater 3	IPA %	IPA Average %
1. A SHOCK	1. general	Yes	Yes	Yes	100	93.2
	2. authentic	Yes	Yes	Yes	100	
	3. accurate	Yes	-	Yes	100	
	4. appropriate	Yes	Yes	-	100	
	5. interesting	Yes	Yes	No	66	
2. DEEP BREATH-ING	1. general	Yes	Yes	Yes	100	100
	2. authentic	Yes	Yes	Yes	100	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	Yes	Yes	Yes	100	
	5. interesting	Yes	Yes	Yes	100	
3. AUTUMN	1. general	Yes	Yes	Yes	100	66.6
	2. authentic	Yes	Yes	Yes	100	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	No	No	-	0	
	5. interesting	Yes	No	No	33	
4. LAUGHTER	1. general	Yes	Yes	Yes	100	100
	2. authentic	Yes	Yes	Yes	100	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	Yes	Yes	Yes	100	
	5. interesting	Yes	Yes	Yes	100	
5. TAKING PICTURES	1. general	Yes	Yes	Yes	100	63.2
	2. authentic	Yes	No	Yes	66	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	No	No	-	0	
	5. interesting	Yes	No	-	50	
6. FISH AND WHALES	1. general	Yes	Yes	Yes	100	80.0
	2. authentic	Yes	Yes	-	100	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	Yes	No	-	50	
	5. interesting	Yes	No	-	50	
7. INTELLI-GENCE PILLS	1. general	Yes	Yes	Yes	100	100
	2. authentic	Yes	Yes	Yes	100	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	Yes	Yes	Yes	100	
	5. interesting	Yes	Yes	Yes	100	

Table 1 (continued)

International Congruence Agreement (ICA) of passage selection and comments from three raters

Passage	Characteristics of the passages	Rater 1	Rater 2	Rater 3	IPA %	IPA Average %
8. BARBECUE	1. general	Yes	Yes	Yes	100	73.2
	2. authentic	Yes	Yes	Yes	100	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	No	No	Yes	33	
	5. interesting	Yes	No	No	33	
9.MOVIE REVIEWS	1. general	Yes	Yes	Yes	100	100
	2. authentic	Yes	Yes	Yes	100	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	Yes	Yes	-	100	
	5. interesting	Yes	Yes	Yes	100	
10. STUDY IN THE U.S	1. general	Yes	Yes	Yes	100	70
	2. authentic	Yes	Yes	Yes	100	
	3. accurate	Yes	Yes	Yes	100	
	4. appropriate	No	No	-	0	
	5. interesting	Yes	No	-	50	

Comments: The followings are comments on the passages from the three raters.

Rater 1: The first rater marked on Passages 3, 5, 8 and 10 as not appropriate in that Passage 3 was culturally specific, Passage 5 had too many words used infrequently in the kind of context. Passage 8 was culturally specific with terms used in particular contexts and Passage 10 contained proper nouns so it was not appropriate for a test. The rater also raised a point that the number of letters deleted be given as clues to the students otherwise the researcher could not accept or reject acceptable and meaningfully correct responses that were not the exact words the researcher had in mind such as 'this' and 'the.' Moreover, the rater added that some of the items contained expressions that would be easily guessed but might not necessarily test students' language ability. Additionally, the rater expressed that it was difficult to be definite when rating the passages as the level of 'interesting' might vary from very interesting to lightly interesting.

Rater 2: The second rater put stars on Passages 4 and 7 and noted that they indicated a great deal of satisfaction in terms of 'interest' while question marks were put on Passages 1,2, and 9 to signify the rater's fair satisfaction. Other passages were rated on either 'Yes' or 'No.' Additionally, the rater commented that the passage selection should have been based on the students' fields of studies and that the rating scales of 'Yes' and 'No' were inferior. The rater suggested that it be appropriate to employ a 'band scale' along with comments for each item so that the rater could rate more definitely. Besides, the rater was concerned that if all students guessed, the test would become void as, according to her opinion, passages in the C-Test format did not provide enough contexts.

Rater 3: The third rater located question marks on some items in Passages 1, 3, 5, 6, 8, 9, and 10 and left them blanks. The rater raised questions on how to mark on the tests and whether misspelling words were counted. Additionally, the rater expressed that personally the rater did not believe in this kind of test and that the way parts of the words were deleted was quite disruptive and unnatural. The rater included that in real life there were plenty of contexts which facilitated understanding and that this kind of test deprived readers of contexts, suggesting that it might work better with native speakers. Also, the rater raised the question whether there were any criteria for text selection and suggested that the researcher describe the procedures in the methodology section. The rater added that it was difficult to validate the passages as the criteria given might make different interpretations among the raters. However, the rater concluded that it would be interesting to see the findings from this research study.

Appendix F

Passages in full texts and in C-Test format in Pilot Study 2 and Main Study

Passage 1 in Pilot Study 2 and Main Study

Passage in full texts.	Passage in C-Test format
1. A SHOCK What you don't know can hurt you. My mother was a model of healthy living. She never smoked. She exercised, ate properly, kept her weight down and her blood pressure under control. She looked and felt great – which is why her heart attacked shocked us all. My mother's experience was not unique. Heart disease is the leading cause of death among American women; it's also tougher to diagnose in women than in men.	1. A SHOCK What you don't know can hurt you. My mot___ was a mo___ of hea____ living. S__ never smo___. She exer_____, ate prop____, kept h__ weight do__ and h__ blood pres____ under con____. She loo___ and fe__ great -- wh___ is w__ her he___ attacked sho____ us a__. My moth____ experience was not unique. Heart disease is the leading cause of death among American women; it's also tougher to diagnose in women than in men.

Passage 2 in Pilot Study 2 and Main Study

Passage in full texts.	Passage in C-Test format
2. DEEP BREATHING You already know how to breathe? You do it every day without even thinking about it. Your breathing technique is not as healthy as you might think. Most of us breathe too shallowly and too quickly. Our lungs and heart would greatly prefer longer, slower, deeper breaths. This is true for general health, and it is true for managing stress. Deep breathing helps drive away the reaction when we're stressed. It sends a signal to your brain to slow down, which results in hormonal and physiological changes that slow heart rate and lower blood pressure.	2. DEEP BREATHING You already know how to breathe? You d_ it ev___ day wit____ even thin____ about i_. Your brea_____ technique i_ not a_ healthy a_ you mi___ think. Mo__ of u_ breathe t__ shallowly a__ too qui____. Our lu___ and he___ would gre____ prefer lon___, slower, dee___ breaths. This is true for general health, and it is true for managing stress. Deep breathing helps drive away the reaction when we're stressed. It sends a signal to your brain to slow down, which results in hormonal and physiological changes that slow heart rate and lower blood pressure.

Passage 3 in Pilot Study 2 and Main Study

Passage in full texts.	Passage in C-Test format
3. LAUGHTER What is the greatest reward of being alive? Is it chocolate, ice cream, vacations, a perfect night's sleep, or the satisfaction of a job well done? Ask a thousand people and you'll get as many answers. But one true pleasure that covers all people is laughter. No matter your age, wealth, race, or living situation, life is good when laughter is frequent. Life is also healthier. Research finds that humor can help you cope better with pain, enhance your immune system, reduce stress, even help you live longer.	3. LAUGHTER What is the greatest reward of being alive? Is i_ chocolate, i__ cream, vaca_____, a per____ night's sl___, or t__ satisfaction o_ a j__ well do__? Ask a thou____ people a__ you'll g__ as ma__ answers. B__ one tr__ pleasure th__ covers a__ people i_ laughter. N_ matter yo__ age, wealth, race, or living situation, life is good when laughter is frequent. Life is also healthier. Research finds that humor can help you cope better with pain, enhance your immune system, reduce stress, even help you live longer.

Passage 4 in Pilot Study 2 and Main Study

Passage in full texts.	Passage in C-Test format
4. INTELLIGENCE PILLS Some scientists have predicted that healthy adults and children may one day take drugs to improve their intelligence. A research group has suggested that such drugs might become as common as coffee or tea within the next couple of decades. In the future, students taking exams might have to take drugs tests like athletes. There are already drugs that are known to improve mental performance, like Ritalin, which is given to children with problems concentrating. A drug given to people who have trouble sleeping also helps people remember numbers.	4. INTELLIGENCE PILLS Some scientists have predicted that healthy adults and children may one day take drugs to improve their intelligence. A rese____ group h__ suggested th__ such dr___ might bec___ as com___ as cof___ or t__ within t__ next cou___ of dec____. In t__ future, stud____ taking ex___ might ha__ to ta__ drugs te___ like athl____. There a__ already dr___ that are known to improve mental performance, like Ritalin, which is given to children with problems concentrating. A drug given to people who have trouble sleeping also helps people remember numbers.

Passage 5 in Pilot Study 2 and Main Study

Passage in full texts.	Passage in C-Test format
5. MOVIE REVIEWS Movie reviews are written, not to tell you what to see, but to help you decide whether or not you would like to see a certain film. Movie critics offer their opinions on the qualities of a particular movie. Sometimes their reviews are positive and other times they are negative. Different reviewers may have very different opinions about the same movie. What one movie reviewer finds entertaining, another may find disappointing.	5. MOVIE REVIEWS Movie reviews are written, not to tell you what to see, but to help you decide whether or not you would like to see a certain film. Movie cri____ offer th___ opinions o_ the qual_____ of a parti_____ movie. Some_____ their rev____ are posi____ and ot___ times th__ are nega____. Different revi_____ may ha__ very diff_____ opinions ab___ the sa__ movie. Wh__ one mo___ reviewer fi___ entertaining, ano____ may find disappointing.

Appendix G

Scores of QPT in normalized T-scores

Table 2

Scores of QPT in normalized T-scores

points	f	F	F - 0.5f	Pr	T- scores
25	1	134	133.5	99.627	77
22	4	133	131	97.761	70
21	10	129	124	92.537	64
20	11	119	113.5	84.701	60
19	8	108	104	77.612	58
18	14	100	93	69.403	55
17	15	86	78.5	58.582	52
16	12	71	65	48.507	50
15	14	59	52	38.806	47
14	11	45	39.5	29.478	45
13	18	34	25	18.657	41
12	9	16	11.5	8.582	36
11	3	7	5.5	4.104	33
10	2	4	3	2.239	30
9	2	2	1	0.746	26

Appendix H

Scores of WBCT in normalized T-scores

Table 3

Scores of WBCT in normalized T-scores

points	f	F	F - 0.5f	Pr	T-scores
66	1	134	133.5	99.627	77
64	1	133	132.5	98.881	73
62	1	132	131.5	98.134	71
59	1	131	130.5	97.388	69
58	2	130	129	96.269	68
56	1	128	127.5	95.149	67
55	2	127	126	94.030	66
54	2	125	124	92.537	65
53	2	123	122	91.045	64
52	1	121	120.5	89.925	63
51	1	120	119.5	89.179	63
50	3	119	117.5	87.687	62
49	1	116	115.5	86.194	61
48	5	115	112.5	83.955	60
46	1	110	109.5	81.716	59
45	1	109	108.5	80.970	58
44	3	108	106.5	79.478	58
43	3	105	103.5	77.239	58
42	5	102	99.5	74.254	57
41	3	97	95.5	71.269	56
40	4	94	92	68.657	55
39	6	90	87	64.925	54
38	7	84	80.5	60.075	53
37	4	77	75	55.970	52

Table 3 (continued)

Scores of WBCT in normalized T-scores

points	f	F	F - 0.5f	Pr	T-scores
36	5	73	70.5	52.612	51
35	1	68	67.5	50.373	50
34	4	67	65	48.507	49
33	3	63	61.5	45.896	49
32	5	60	57.5	42.910	48
31	4	55	53	39.552	47
30	1	51	50.5	37.687	47
29	6	50	47	35.075	46
28	1	44	43.5	32.463	45
27	3	43	41.5	30.970	45
25	3	40	38.5	28.731	44
24	6	37	34	25.373	43
23	2	31	30	22.388	42
22	4	29	27	20.149	42
21	3	25	23.5	17.537	41
20	4	22	20	14.925	40
18	1	18	17.5	13.060	39
17	1	17	16.5	12.313	38
16	1	16	15.5	11.567	38
15	2	15	14	10.448	37
14	1	13	12.5	9.328	37
13	7	12	8.5	6.343	35
12	1	5	4.5	3.358	32
10	2	4	3	2.239	30
7	1	2	1.5	1.119	27
6	1	1	0.5	0.373	23

Appendix I

Scores of QPT40, WBCT100, WBCT40 and WBCT Level

Table 4

Scores of QPT40, WBCT100, WBCT40 and WBCT Level

Students' ID	QPT 40	WBCT 100	WBCT 40	WBCT Level
1	21	58	23.2	1
2	17	54	21.6	1
3	13	21	8.4	0
4	13	34	13.6	0
5	16	37	14.8	0
6	13	10	4	0
7	18	29	11.6	0
8	13	24	9.6	0
9	14	38	15.2	0
10	14	23	9.2	0
11	17	13	5.2	0
12	16	38	15.2	0
13	21	50	20	1
14	12	29	11.6	0
15	12	39	15.6	0
16	13	52	20.8	1
17	13	13	5.2	0
18	14	29	11.6	0
19	21	27	10.8	0
20	21	55	22	1
21	17	36	14.4	0
22	13	29	11.6	0
23	18	31	12.4	0

Table 4 (continued 1)

Scores of QPT40, WBCT100, WBCT40 and WBCT Level

Students' ID	QPT 40	WBCT 100	WBCT 40	WBCT Level
24	20	39	15.6	0
25	13	6	2.4	0
26	15	37	14.8	0
27	13	36	14.4	0
28	18	40	16	1
29	12	42	16.8	1
30	14	30	12	0
31	21	46	18.4	1
32	14	31	12.4	0
33	21	40	16	1
34	14	16	6.4	0
35	19	44	17.6	1
36	15	38	15.2	0
37	18	50	20	1
38	18	55	22	1
39	15	25	10	0
40	19	22	8.8	0
41	22	36	14.4	0
42	20	34	13.6	0
43	11	17	6.8	0
44	16	38	15.2	0
45	17	20	8	0
46	22	54	21.6	1
47	14	20	8	0
48	16	43	17.2	1

Table 4 (continued 2)

Scores of QPT40, WBCT100, WBCT40 and WBCT Level

Students' ID	QPT 40	WBCT 100	WBCT 40	WBCT Level
49	15	59	23.6	1
50	18	29	11.6	0
51	13	32	12.8	0
52	17	48	19.2	1
53	16	33	13.2	0
54	15	36	14.4	0
55	12	12	4.8	0
56	15	38	15.2	0
57	18	51	20.4	1
58	17	58	23.2	1
59	16	13	5.2	0
60	19	33	13.2	0
61	17	22	8.8	0
62	17	14	5.6	0
63	16	37	14.8	0
64	19	39	15.6	0
65	20	44	17.6	1
66	21	64	25.6	2
67	20	66	26.4	2
68	15	24	9.6	0
69	11	40	16	1
70	17	41	16.4	1
71	17	25	10	0
72	13	21	8.4	0
73	16	53	21.2	1
74	17	36	14.4	0
75	22	53	21.2	1

Table 4 (continued 3)

Scores of QPT40, WBCT100, WBCT40 and WBCT Llevel

Students' ID	QPT 40	WBCT 100	WBCT 40	WBCT Level
76	18	15	6	0
77	15	27	10.8	0
78	18	13	5.2	0
79	13	20	8	0
80	15	41	16.4	1
81	14	22	8.8	0
82	18	50	20	1
83	12	24	9.6	0
84	10	38	15.2	0
85	18	48	19.2	1
86	20	40	16	1
87	13	29	11.6	0
88	17	49	19.6	1
89	19	23	9.2	0
90	13	34	13.6	0
91	16	18	7.2	0
92	14	32	12.8	0
93	17	27	10.8	0
94	21	32	12.8	0
95	16	22	8.8	0
96	15	32	12.8	0
97	9	7	2.8	0
98	19	48	19.2	1
99	16	31	12.4	0
100	13	20	8	0
101	19	13	5.2	0
102	15	39	15.6	0

Table 4 (continued 4)

Scores of QPT40, WBCT100, WBCT40 and WBCT Level

Students ID	QPT 40	WBCT 100	WBCT 40	WBCT Level
103	18	41	16.4	1
104	11	13	5.2	0
105	20	45	18	1
106	13	42	16.8	1
107	15	42	16.8	1
108	12	42	16.8	1
109	17	39	15.6	0
110	12	38	15.2	0
111	10	39	15.6	0
112	13	37	14.8	0
113	9	34	13.6	0
114	20	48	19.2	1
115	18	44	17.6	1
116	16	35	14	0
117	21	62	24.8	2
118	25	25	10	0
119	20	10	4	0
120	13	43	17.2	1
121	20	24	9.6	0
122	20	48	19.2	1
123	14	33	13.2	0
124	12	24	9.6	0
125	17	24	9.6	0
126	12	28	11.2	0
127	15	13	5.2	0
128	18	31	12.4	0
129	14	42	16.8	1

Table 4 (continued 5)

Scores of QPT40, WBCT100, WBCT40 and WBCT Level

Students ID	QPT 40	WBCT 100	WBCT 40	WBCT Level
130	20	32	12.8	0
131	15	43	17.2	1
132	21	15	6	0
133	22	56	22.4	1
134	19	21	8.4	0

BIOGRAPHY

Naddhaporn (Dhirasedh) Kammasorn is an Assistant Professor of Rajamangala University of Technology Tawan-ok, in the Faculty of Liberal Arts, English Department, at Chakrabongse Bhuvanath Campus in Bangkok. She received a Master's Degree of Education in Foundations of Education from Chulalongkorn University in 1987, and a Master's Degree of Arts in Language and Communication from National Institute of Development Administration (NIDA) in 2004. Her research interests are on language assessments and the uses of IT in language learning and evaluation. She can be contacted through Chakrabongse Bhuvanath Campus, 122 Vipavadee-Rangsit Rd., Dindaeng, Bangkok 10400, THAILAND, or via email at pnadd@yahoo.com.