QSAT: The Web-Based mC-Test as an Alternative English Proficiency Test

Somsak Boonsathorn*
Mae Fah Luang University, Thailand

Chayaporn Kaoropthai
Mae Fah Luang University, Thailand

Abstract

Reversing the original C-Test (Raatz & Klein-Braley, 1981), the modified C-Test (mC-Test) deletes the *first* half (instead of the second half) of every second word (Boonsathorn, 1987). It is also known as the X-Test (Cleary, 1988; Prapphal, 1994, 1996). This web-based version was explored to investigate its reliability, face validity, criterion-related validity, and concordance with the Quick Placement Test (QPT). The participants were 585 undergraduate students studying varying academic areas, in a government university in northern Thailand during the 2010 academic year; all the students were volunteers. The instruments used included: (1) the Quick Placement Test (QPT), (2) the Web-Based mC-Test (WB mC-Test), and (3) a questionnaire concerning the face validity of the WB mC-Test. The findings revealed that the WB mC-Test had high reliability and high face validity. The Pearson correlation between the WB mC-Test and the QPT was significant at a medium level. Using a statistical model, it was found that the WB mC-Test could differentiate subjects into 4 levels based on the ALTE (Association of Language Testers in Europe): Level 0 Beginner to Level 3 Upper Intermediate. The WB mC-Test is supported to be a practical preliminary self-assessment test for EFL university students.

Keywords: C-Tests, X-Tests, English proficiency tests, self-assessment tests, web-based testing

Introduction

English proficiency is important when studying at the tertiary level in Thailand especially in a university where English is a medium of instruction. Students admitted into academic programs are generally well-prepared for their academic disciplines. Students' English proficiency has often been a major factor in failure, since the minimum requirement of their O-NET (Ordinary National Educational Test) score in English is only 35%. Mae Fah Luang University, established in 1998, has been the only autonomous state-affiliated university in Thailand which uses English as a medium of instruction in all subject areas except law, students' English proficiency has always been a major factor affecting their academic success or failure. The Self-Access Language Learning Center (SALLC) at Mae Fah Luang has provided necessary facilities and equipment to help further develop students' English proficiency. The problem to date is that many students do not know where to start with in the SALLC. Students often do not know their level of English proficiency, presenting an urgent need for a practical instrument to help students estimate their English proficiency levels. In this study, the QSAT (Quick Self-Assessment Test) was proposed using the Web-Based mC-Test as a preliminary self-assessment test of general English proficiency. The purpose of the present study was to develop and validate a web-based modified C-Test (WB mC-Test) to use as a quick self-assessment test of general English proficiency. To evaluate the proposed WB mC-test, five major aspects of test qualities for this were focused on: reliability, difficulty suitability, criterion-related validity, concordance with the Quick Placement Test score, and face validity.

* *Email: somsakb89@yahoo.com. Tel.: 66-818-05-8282. Address: Mae Fah Luang University, 333 Moo1, Thasud, Muang, Chiang Rai, 57100, Thailand.*

The C-Test

Alderson's (1979, 1980) and Klein-Braley's (1981) critical studies of the systematic every n^{th} word deletion cloze test revealed many significant points of criticism. Some of the major points were about rates and starting points of deletion that affected the difficulty, reliability, and validity of the tests. For example, the use of only one or two cloze passages could be the cause of content bias. There was also a problem with scoring procedures. In addition, the fact that educated native speakers could rarely obtain a perfect score created doubts about acceptability judgments in scoring.

In order to resolve the shortcomings facing the cloze test, Klein-Braley (1984) proposed six criteria for the new format of a reduced redundancy test, the C-Test:

1. The C-Test should use several different texts.
2. It should have at least 100 deletions (items).
3. Adult native speakers should obtain nearly perfect scores.
4. The deletions should affect a representative sample of the text.
5. Only exact-word scoring should be possible.
6. The test should have high reliability and validity. (p.136)

The C-Test is normally comprised of four to six short texts constructed according to what Klein-Braley and Raatz (1984) called *The Rule of 2*. The deletion in each text begins in the second sentence by deleting the second half of every second word until the required number of mutilations is reached, while the rest of the text continues to the end of the paragraph. The following is an example of a C-Test passage from Klein-Braley and Raatz (1984).

The C-Test

> There are usually five men in the crew of a fire engine. One o_ them dri_ _ _ the eng_ _ _. The lea_ _ _ sits bes_ _ _ the dri_ _ _ . The ot_ _ _ firemen s_ _ inside t_ _ cab o_ the fi_ _ engine. T_ _ leader h_ _ usually be_ _ in t_ _ Fire Serv_ _ _ _ for ma_ _ years. H_ will kn_ _ how t_ fight diff_ _ _ _ _ sorts o_ fires. S_ , when t_ _ firemen arr_ _ _ at a fire, it is always the leader who decides how to fight a fire. He tells each fireman what to do. [25 items] (p.136)

Since its initiation by Raatz and Klein-Braley in 1981, the C-Test could be considered the best in the family of tests of reduced redundancy. Many research studies have supported the notion that the C-Tests are theoretically and empirically valid and reliable tests of overall language proficiency (Babaii & Ansary, 2001; Cohen, Segal, & Bar-Siman-To, 1984; Connelly, 1997; Dörnyei & Katona, 1992; Klein-Braley, 1997; Klein-Braley & Stevenson, 1981; Raatz & Klein-Braley, 1981; Rouhani, 2008).

Despite the increasing literature supporting C-Tests, some research studies have questioned the effectiveness C-Tests (Bradshaw, 1990; Cleary, 1988; Weir, 1990). Most of these studies found that the C-Tests were not automatically valid and reliable tests of overall language proficiency. They were often too easy and lacked face validity.

Relating to face validity, Nevo (1985), in a critical study of face validity, concluded that there are two basic viewpoints about face validity. The first one is to separate face validity from other types of validity, while the second considers face validity "an important feature of any psychological or educational test intended for practical use" (p. 287). According to the latter viewpoint, a test of high face validity may have an advantage over others in terms of test takers' motivation, interest, and satisfaction; it can convince users to implement it, and it can also help improve public relations.

The theoretical framework for the C-Test is literally an adoption of every n^{th} word deletion cloze framework. Taylor (1953), the initiator of the cloze test, appeared to be inspired by Gestalt psychology and

information theory as theoretical bases for the cloze procedure. The closure principle in Gestalt psychology contends that individuals can perceive objects such as shapes, letters, pictures, etc., as being whole when they are not complete. Specifically, when parts of a whole picture are missing, our perceptions fill in the visual gap. A familiar example of this concept is our ability to see a broken circle or rectangle as whole by mentally closing the gaps. According to Gestalt psychologists, the process of learning consists of global comprehension first to be followed later by the comprehension of detail (Stansfield & Hansen, 1983). The reason test takers are able to restore mutilated texts can also be supported by information theory, particularly by the concept of redundancy. According to Spolsky (1971), natural language contains redundancy to safeguard a message against noise which may interfere with the message. As Spolsky noted, "messages in normal language can be understood even though a good proportion of them is omitted or masked" (p.167). This concept of reduced redundancy is also known as *expectancy grammar*, as coined by Oller (1976, 1979). Adopting the same theoretical framework with the cloze test, the C-Test is therefore based on Gestalt's closure principle and the reduced redundancy principle (RRP) (Klein-Braley, 1981).

The theoretical rationale for using the C-Test to measure reading comprehension can be explained by psycholinguistic theories of the reading process. One of the most widely accepted reading models to describe the reading strategies used by readers is that of Goodman (1967). Goodman proposed that readers use graphophonic, syntactic, and semantic information as they engage in the reading act, and that the best readers use the least amount of text information available. Smith (2012) argued that reading as an activity involves two forms of information: the visual (what is on the printed page), and the nonvisual (the reader's language competence and their background experiences). The reader uses these two forms of information to understand what the author is describing. Therefore, the more nonvisual information one has, the less they need to rely on visual information. Although Goodman and Smith used different explanations, they essentially agreed that an efficient reader usually uses a minimum amount of text or visual information.

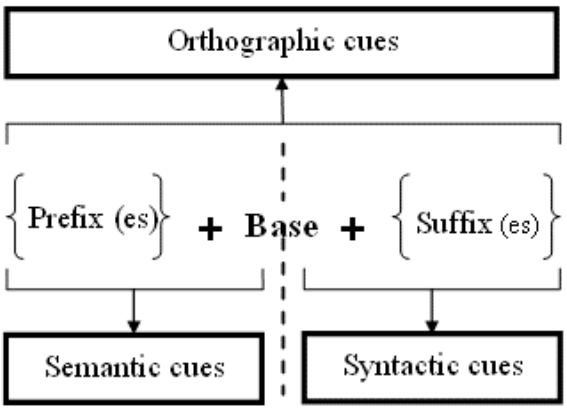


Figure 1. English Word Formation Based on Goodman's Reading Model (1967).

The mC-Test

Based on the Goodman Model, when the second half of every second word is deleted in the C-Test, some graphophonic/orthographic cues are still present in each item. Removing half of the word leaves the reader with a fair amount of information even if the deletion occurs in every second word. A challenging question was then, between the semantic cues (present in the first half of a word) and the syntactic cues (present in the second half of a word), which type of cues would prove to require more of the reading strategies used in normal reading? Relating to the English word formation above, the theoretical framework for proposing the mC-Test can be outlined diagrammatically as follows:

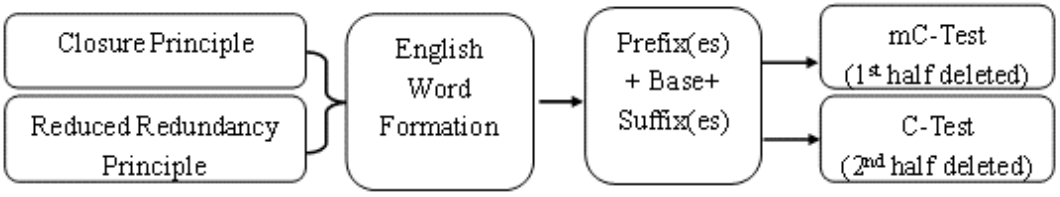


Figure 2. Theoretical Framework of the C-Test and mC-Test

Figure 2 clearly depicts the difference between the main aspect of an mC-Test item and a C-Test item. While the semantic cues are present in the C-Test item, only syntactic cues are normally available in the mC-Test item. Consequently, the mC-Test will be more difficult and necessitate readers to employ more language skills and strategies to figure out the word than those required for the C-Test.

The author (Boonsathorn, 1987) initiated this new format, the mC-Test, to resolve the problems facing the conventional C-Test. In the mC-Test, the first half, instead of the second half, of every second word is deleted. The results revealed that the C-Tests and the mC-Tests were highly reliable and valid as overall language proficiency tests for both native and non-native learners and the mC-Test was shown to be more difficult and had better discrimination power than the C-Test. Cleary (1988) used the same procedure of *left-hand deletions* and coined the term *X-Test*. Some research studies (Prapphal, 1994, 1996; Sigott & Kobrel, 1993) as well as the author's student researchers were able to demonstrate that the mC-Tests or X-Tests are valid and reliable overall language proficiency tests which are more difficult than the C-Tests. The following is an example of the mC-Test constructed from the same text used in the sample C-Test.

> The mC-Test
>
> There are usually five men in the crew of a fire engine. One _f them _ _ _ ves the _ _ _ ine. The _ _ _ der sits _ _ _ ide the _ _ _ ver. The _ _ _ er firemen _ _ t inside _ _e cab _ f the _ _ re engine. _ _ e leader _ _ s usually _ _ en in _ _ e Fire _ _ _ vices for _ _ ny years. _ e will _ _ ow how _ o fight _ _ _ _ _ rent sorts _ f fires. _ o, when _ _ e firemen _ _ _ ive at a fire, it is always the leader who decides how to fight a fire. He tells each fireman what to do. [25 items]

Since there have not been any research studies about the mC-Tests that are web-based the author decided to investigate the possibility of using the Web-Based mC-Test as a quick self-assessment test of general English proficiency.

Research Questions

This study aims to address five research questions about the WB mC-Test:

RQ 1: How high is the reliability of the WB mC-Test?
RQ 2: What is the difficulty level of the WB mC-Test?
RQ 3: How high is the criterion-related validity against the QPT?
RQ 4: What is the concordance between the WB mC-Test and the QPT?
RQ 5: How high is the face validity of the WB mC-Test?

Methodology

Participants

Since many participants were necessary, the researcher decided to use a voluntary sample, using 585 undergraduate students who were invited to participate in the study by their English course instructors. The students were aware that their grades would not be affected by their decision to or not to participate in the study. The study was approved by the university research ethics committee. The students who were interested to participate were asked to read and sign the informed consent. Each participant was given 100 Thai baht (approximately US$ 3) as an incentive. These students were first- to third-year students from all majors, enrolled at Mae Fah Luang University in the first semester of academic year 2010.

Instruments

The instruments for this study included two tests and a questionnaire:

1. The Quick Placement Test (QPT), Paper and Pen Test Version 2: A standardized test by the University of Cambridge Local Examinations Syndicate (Syndicate, 2001). The QPT consists of two parts: 40 questions in Part 1 and 20 questions in Part 2; the total possible score for the whole is 60.
2. The Web-Based Modified C-Test (WB mC-Test): An mC-Test using the software program http://clozeonline.us/mindex1.html on the Internet. The WB mC-Test consisted of five short texts with 20 mutilated words in each, totaling 100 items, 100 points. The test was designed in the mC-Test format, where the first half (instead of the second half) of every second word is deleted (see Appendix).
3. The Questionnaire: A closed-ended, Likert-Type questionnaire regarding the participants' opinions about the WB mC-Test. There were three parts in the questionnaire:

   Part 1 asked what the participants thought of "the WB mC-Test measures" and consisted of six items.

   Part 2 asked whether the participants thought "the WB mC-Test is an efficient test of general English proficiency" and consisted of one item.

   Part 3 was an open-ended question that asked the participants to make comments and suggestions.

Data Collection Procedure

Since it was necessary to have a sufficient number of participants to ensure the reliability and validity of the test, test administrations were conducted three times to obtain data from 585 participants. In all three test administrations, the same procedures were adopted. The participants started by practicing the *sample* WB mC-Test for 10 minutes before doing the *official* WB mC-Test and completing the questionnaire within 35 minutes. After a five-minute break, the participants took the QPT, which lasted 30 minutes.

Analytical Procedure

The data collected were processed through statistical procedures to arrive at answers to the above research questions. A pilot study was conducted to assess the readability suitability of test texts, using the Microsoft Word program to obtain Flesch-Kincaid readability grade levels. To answer the first research question, Cronbach's alpha reliability was performed. The mean scores and percentage points were calculated to answer RQ2. To answer RQ3, Pearson's correlation coefficient and overlapping variance were used to establish the criterion-related validity. To propose a concordance chart, answering RQ4, a test of regression was employed to investigate the degree of prediction of the WB mC-Test compared to the QPT so as to obtain a formula to predict the QPT scores. Finally, the data from a four-point Likert's scale questionnaire were analyzed, using means and percentage, to answer RQ5.

Pilot study

The pilot study of the WB mC-Test was conducted with 35 undergraduate students at Mae Fah Luang University at the beginning of the first semester of academic year 2010. The purpose of the pilot study was to assess the readability suitability and the readability level of the test texts chosen. Microsoft Word was employed to calculate Flesch-Kincaid readability grade levels. Flesch-Kincaid readability grade levels are generally used to determine the difficulty level of text appropriate for each school grade level of English native students. For example, readability grade 9 will be appropriate for 9 graders and readability grade 13 for first-year university students. The results revealed that one of the texts was inappropriate because its grade level was 6.7, which was too easy for the target sample; the texts chosen for the main study were as follows:

Table 1
Texts Used in the WB mC-Test Main Study

Topic	No. of words	Readability
Car	86	7.8
Diet	83	8.7
Blood	101	9.2
Behavior	110	9.4
Computer	90	12.8
Mean (*N*=5)	94	9.6

Table 1 shows that the texts used for the main study ranged from 83 to 110 words long and their readability grade levels were between 7.8 to 12.8, averaging 9.6. The results reflect that the text is suitable for Grade 10 (English native students) and should be appropriate for EFL first-year university students. For testing purposes, texts with varied readability levels were used.

Results

RQ 1: How high is the reliability of the WB mC-Test?

Table 2
Reliability

	No. of items	Alpha
Text 1	20	.85
Text 2	20	.72
Text 3	20	.81
Text 4	20	.64
Text 5	20	.69
Total	100	.89

The statistics in Table 2 show Cronbach's alpha reliability of the mC-Test. Although the Alphas for Texts 4 and 5 are not very high, the overall reliability is .89, which is considered high (Lazaraton&Hatch, 1991).

RQ 2: What is the difficulty level of the WB mC-Test?

The descriptive statistics in Table 3 shows that the mean of all 585 participants for the QPT is 24.412 (out of the total 60, or 40.687 %) and the WB mC-Test as 52.015 (out of the total 100, or 52.015%). The standard deviation reveals that there are moderate variances in both test scores. Since the mean score of the WB mC-Test is around 50%, it is likely not too difficult nor too easy for this target sample, thus the difficulty level is considered

appropriate. This result implies that the readability grade level of the texts used for the WB mC-Test can range distributively from readability grade levels 8 to 13 approximately (see Table 1).

Table 3
Descriptive Statistics of the QPT and the WB mC-Test Results

	N	Minimum	Maximum	Mean	S.D
QPT	585	9.0	44.0	24.412	5.505
Web	585	2.0	86.0	52.015	15.199

RQ 3: How high is the criterion-related validity against the QPT?

Table 4
Paired Mean Scores of the WB mC-Test and the QPT (adjusted)

	N	Mean	S.D.	Std. Error Mean	t value	Sig.(2-tailed)
Web	585	52.015	15.198	.628	19.844	.000***
QPTadj	585	40.687	9.175	.379		

*** p< .001

Table 4 displays the paired mean scores of the two tests, using the adjusted QPT mean score out of the total, 100. The mean scores of the two tests are significantly different at the .001 level. Although the QPT is more difficult than the WB mC-Test, the mean scores of 40.687 and 52.015 are still in the range of ±50, which is appropriate for this target group.

Table 5
Correlation between the WB mC-Test and the QPT

	Web	QPTadj	Sig.(2-tailed)
Web	1	.446***	.000***
QPTadj	.446***	1	

*** p< .001

Table 5 shows that there is a significant correlation between the WB mC-Test and the QPT. The correlation coefficient (r) value is .446 (p< .001), which is considered to be moderate. The overlapping varience (r^2) value of .199 reveals that the two tests are measuring different areas of students' language skills, with only a 19.9% overlap (Bailey, 1998). The criterion-related validity of the WB mC-test is thus rather low.

RQ 4: What is the concordance between the WB mC-Test and the QPT?

Table 6 displays the concordance between the WB mC-test scores and the QPT scores. The concordance test helps estimate the relationship of the two test score. Since the maximum QPT scores of participants was only 44 (see Table 3), the WB mC-Test could probably predict up to the Upper Intermediate level for those who score 90 or more. The scatterplot in Figure 3 depicts the relationship between the WB mC-Test scores and the QPT scores based on the regression equation used.

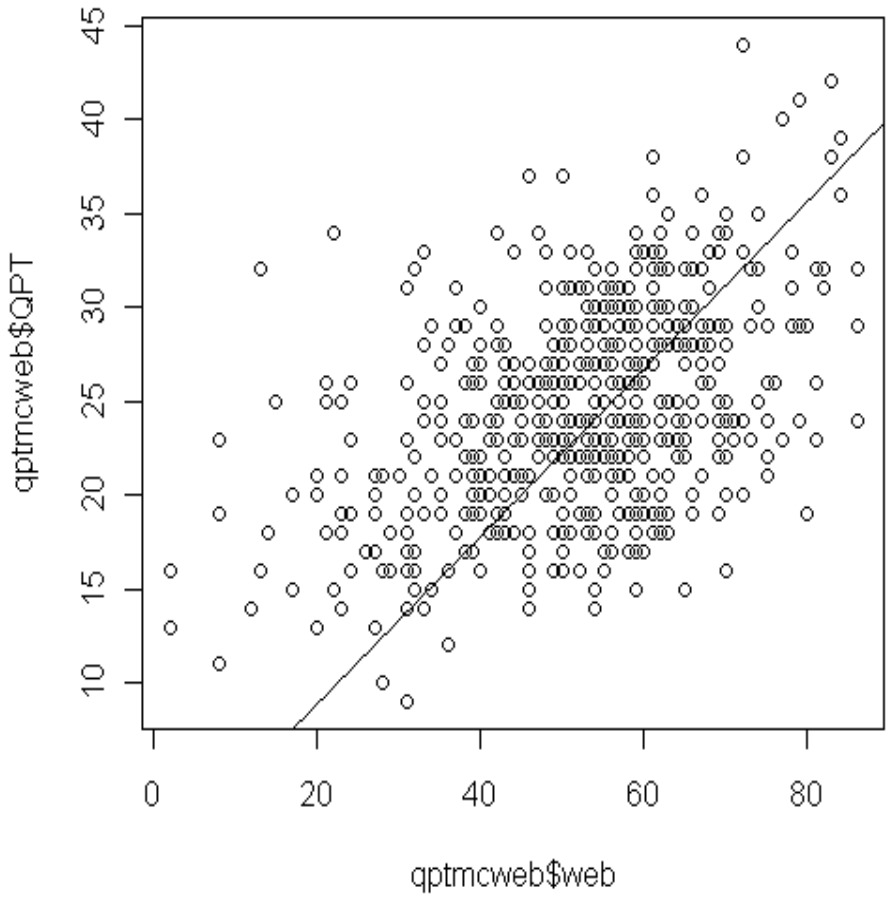


Figure 3. Scatterplot of Predicted QPT Scores Versus WB mC-Test Scores

Table 6
WB mC-Test and Predicted QPT Scores

Parameter Estimates Dependent Variable: QPT							
Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
web	0.4452	0.0051	87.5527	0.0000	0.4352	0.4551	
				Lower Bound	Upper Bound		
#0 -17 Beginner	0	17		-	38.19		max 100
#18-29 Elementary	18	29		40.43	65.14		
#30 - 39 Lower Intermediate	30	39		67.39	87.61		
#40 - 47 Upper Intermediate	40	47		89.85	105.58		
#48 - 54 Advanced	48	54		107.83	121.30		
#55 - 60 Very Advanced	55	60		123.55	134.78		

Table 7
Concordance Chart

	QPT		WB MC-Test	
			Lower Bound	Upper Bound
#0 -17 Beginner	0	17	0	38.19
#18-29 Elementary	18	29	40.43	65.14
#30-39 Lower Intermediate	30	39	67.39	87.61
#40-47 Upper Intermediate	40	47	89.85	100.00
#48-54 Advanced	48	54		
#55-60 Very Advanced	55	60		

Based on a predictive statistical model (see Table 6 & Figure 3), it was found that the WB mC-Test can differentiate participants into four levels of the ALTE (Syndicate, 2001), from Beginner to Upper Intermediate. There were no participants advanced enough to reach a QPT score appropriate for the Advanced or Very Advanced level.

Table 8
Number of Participants in ALTE levels

qptcat * webnointcat Crosstabulation
Count

		webnointcat			Total
		Beginner	Elementary	lower intermediate	
qptcat	Beginner	33	22	1	56
	Elementary	87	294	50	431
	lower intermediate	10	59	24	93
	Upper intermediate	0	0	5	5
	TOTAL	130	375	80	585

Table 8 reveals that based on the concordance chart in Table 7, most participants (431) were at the Elementary level, with a WB mC-Test score between 40 and 65. Fifty-six participants were at Level 0, Beginner; while 93 were at Level 2, Lower Intermediate; and only five were at Level 3, Upper Intermediate.

RQ 5: How high is the face validity of the WB mC-Test?

Table 9
Face Validity of the WB mC-Test

Questionnaire						
1. The Web - Based mC-Test measures:	N	4	3	2	1	M
1.1 Vocabulary and Spelling	589	147 (25.0%)	387 (65.7%)	47 (8.0%)	8 (1.4%)	3.14
1.2 Grammar	588	126 (21.4%)	378 (64.3%)	77 (13.1%)	7 (1.2%)	3.06
1.3 Analytical Ability	585	128 (21.9%)	363 (62.1%)	84 (14.3%)	10 (1.7%)	3.04
1.4 Background Knowledge	588	123 (20.9%)	338 (57.5%)	117 (19.9%)	10 (1.7%)	2.98
1.5 Reading Comprehension	584	139 (23.8%)	364 (62.3%)	75 (12.9%)	6 (1.0%)	3.09
1.6 General English Knowledge	584	150 (25.7%)	357 (61.1%)	70 (12.0%)	7 (1.2%)	3.11
2. The Web - Based mC-Test is an efficient test of general English proficiency.	552	130 (23.6%)	342 (61.9%)	70 (12.7%)	10 (1.8%)	3.07

Note. 4 = strongly agree; 3 = agree; 2 = disagree; 1 = strongly disagree

To determine the degree of agreement and disagreement, the following intervals of mean scores were adopted.

1.00-1.50	=	strongly disagree/very low
1.51-2.50	=	disagree/low
2.51-3.50	=	agree/high
3.51-4.00	=	strongly agree/very high

The data from the questionnaire indicated that the mean scores concerning the characteristics of the WB mC-Test ranged from 2.98 to 3.14. These mean scores could be interpreted as participants agreeing that the WB mC-Test had all of those aspects or attributes. The data revealed that the highest mean score was on Vocabulary and Spelling (3.14), while the lowest was on Background Knowledge (2.98).

From the mean scores, it can be concluded that the participants agreed that "The WB mC-Test measures, Vocabulary and Spelling, General English Knowledge, Reading Comprehension, Grammar, Analytical Ability, and Background Knowledge," respectively. They also agreed (with the test) that "The WB mC-Test is an efficient test of general English proficiency" (M=3.07). Statistically, 85.5% of the participants either agreed or strongly agreed that "The Web-Based mC-Test is an efficient test of general English proficiency."

In order to investigate the perceptions of expert participants, six English instructors (five Thais and one native speaker of English) with a minimum of five-years of EFL teaching experience at the tertiary level, were also asked to do the mC-Test and complete the questionnaire. All six participants rated either *agree* or *strongly agree* for four of the six aspects: *1.1 Vocabulary and Spelling*, *1.3 Analytical Ability*, *1.5 Reading Comprehension*, and *1.6 General English Knowledge*. For *1.2 Grammar* and *1.4 Background Knowledge*, five participants rated either *agree* or *strongly agree*, while one rated *disagree* on each. Interestingly, four participants agreed (3 *agree* and 1 *strongly agree*) that "*The mC-test is an efficient test of general English proficiency*," whereas two chose *disagree*; both of these participants, however, agreed that the mC-Test measures various aspects of the language. The one native speaker of English disagreed because "it tests advanced levels of English proficiency, not general English proficiency"; thus, it might depend on how one defines the scope of general English proficiency.

Conclusion and Discussions

Evidence from statistical analyses supported the notion that overall the WB mC-Test has high reliability, which is an important quality for test usefulness (Bachman & Palmer, 1996). The overall Cronbach's alpha was .89, which was consistent with previous studies about mC-Tests (Boonsathorn, 1987; Prapphal, 1994, 1996; Rungruangthum, 2005; Wonghiransombat, 2013). The mean score of the WB mC-Test was 52.015%, which supported the argument that readability grade levels between 7.8 and 12.8 were suitable for Mae Fah Luang University undergraduate students. The results were in line with Kammasorn's (2008) study, which found that for relatively low proficiency undergraduate students, the readability grade levels between 5.6 and 11.3 were appropriate for their Web-Based C-Test.

The correlation between the WB mC-Test and the QPT was .446 (r^2 = .199). The correlation was significant at the .001 level, but not very high. The criterion-related validity of the WB mC-test against the QPT is thus rather low. Kammasorn's (2008) findings also yielded a significant but not high correlation between the web-based C-Test and the QPT (r = .340; r^2 = .116). The moderate correlation may imply that both the C-Test and the mC-Test focus on different areas within English from the QPT. Even though the two types of test purport to be general English proficiency tests, they differ in certain aspects. The QPT, like most standardized proficiency tests, assesses three specific areas of English—reading, vocabulary, and grammar. The C-test and the mC-Test, on the other hand, are integrative tests, requiring test-takers to incorporate different skills or abilities in completing the test. The QPT consists of two parts: Part 1 is taken by all students and Part 2 is for higher ability students only (Syndicate, 2001). The C-Test and mC-Test usually consist of four to six short passages, assessing test-takers' overall abilities. The QPT is a multiple-choice test, whereas the C-Test and mC-Test require test-takers to fill in the missing parts of words. For these reasons, it can be concluded that the Web-Based mC-Test does not conclusively measure the same types of abilities as the QPT.

The findings reveal that the WB mC-Test could differentiate students into four levels of ALTE, from Level 0 (Beginner) to Level 3 (Upper Intermediate). The Concordance Chart shows that there were no participants advanced enough to reach a high QPT score and only five participants were at the Upper Intermediate level. Kammasorn (2008) had a similar problem in their study because the subjects' QPT scores were between Level 0 (Beginner) and Level 1 (Elementary), enabling the Web-Based C-Test to differentiate into only two levels.

In terms of face validity, the WB mC-Test appears to be quite high, with a mean score from 2.98 to 3.14, which is the *agree* level. To be precise, 85.5% of the participants either agreed or strongly agreed that the WB mC-Test is an efficient test of general English proficiency. The results were consistent with the researcher's (Boonsathorn, 1987, 2000, 2007; Boonsathorn, Getkham, & Boonsathorn, 2007) studies and Kammasorn's (2008) study, while scholars such as Weir (1990) and Bradshaw (1990) found that the C-Test had low face validity.

Interestingly, Oller and Conrad (1971) pointed out that in Taylor's (1953) pioneering cloze study with the aim of measuring L1 subjects' reading comprehension, it was found that cloze scores also correlated with other attributes including vocabulary knowledge and IQ. In their experimental study, Oller and Conrad (1971) found that their cloze test discriminated well between beginning, intermediate, and advanced ESL students as well as English L1 speaking TEFL graduates. In addition, the study revealed a high multiple correlation of .88 with the UCLA ESLPE test and considerable correlations with the subtests: Grammar (.58), Vocabulary (.59), Reading (.80), and Dictation (.82). Oller and Conrad concluded that "with further experimentation and refinement, the cloze method may play an extremely useful role in the placement of non-native speakers of English and in the diagnosis of their special language problems" (p. 192).

The findings of the present study regarding face validity appear to lend support to the above mentioned empirical studies. Specifically, vocabulary and spelling were rated the highest rank of attributes (M= 3.14), followed by general English knowledge (M= 3.06), analytical ability (M= 3.04), and background knowledge (M= 2.98), respectively. These mean scores are considered high and were all at the level of *agree*. The students' and experts' judgment data appear to support that the test takers agreed that the WB mC-Test measured all these attributes. For the question, as to whether The Web-Based mC-Test is an efficient test of general English proficiency, the majority of students (85.5%) either agreed or strongly agreed that it was (M= 3.07).

Although it may not be able to lend theoretical support to the construct validity of the test, face validity is a characteristic of tests that can be measured (Nevo, 1985). Therefore, students' and experts' judgment data can partially support other characteristics of the test, especially practicality.

Despite some limitations, this study has supported the promise of the WB mC-Test as a possible alternative English proficiency test—a practical preliminary self-assessment tool for university students; test takers can receive instant score reports as soon as they complete the test. For future research, an investigation of washback, the effects on teaching and learning, of the WB mC-Test should be performed. Another recommendation, which is important, but has not been investigated here, is validating the WB mC-Test empirically by correlating it with other specific attributes of language proficiency to help confirm their connections.

References

Alderson, J. C. (1979). The Cloze Procedure and Proficiency in English as a Foreign Language. *TESOL Quarterly, 13*(2), 219-227. doi:10.2307/3586211

Alderson, J. C. (1980). Native and Nonnative Speaker Performance on Cloze Test 1. *Language Learning, 30*(1), 59-76. doi:10.1111/j.1467-1770.1980.tb00151.x

Babaii, E., & Ansary, H. (2001). The C-test: a valid operationalization of reduced redundancy principle? *System, 29*(2), 209-219. doi:http://dx.doi.org/10.1016/S0346-251X(01)00012-4

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford: OUP.

Bailey, K.M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. Cambridge, MA: Heinle&Heinle.

Boonsathorn, S. (1987). *C-Tests, proficiency, and reading strategies in ESL*: National Library of Canada.

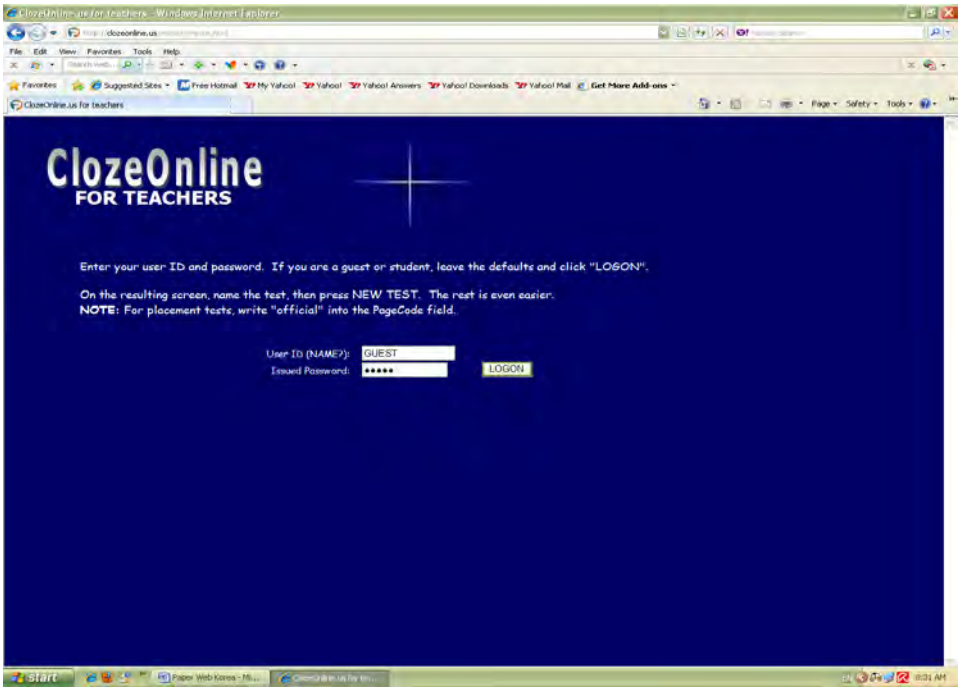
Boonsathorn, S. (2000). S-Tests: A rational C-Test for advanced students. *Thai Journal of Development Administration, 40*(1), 65-74.

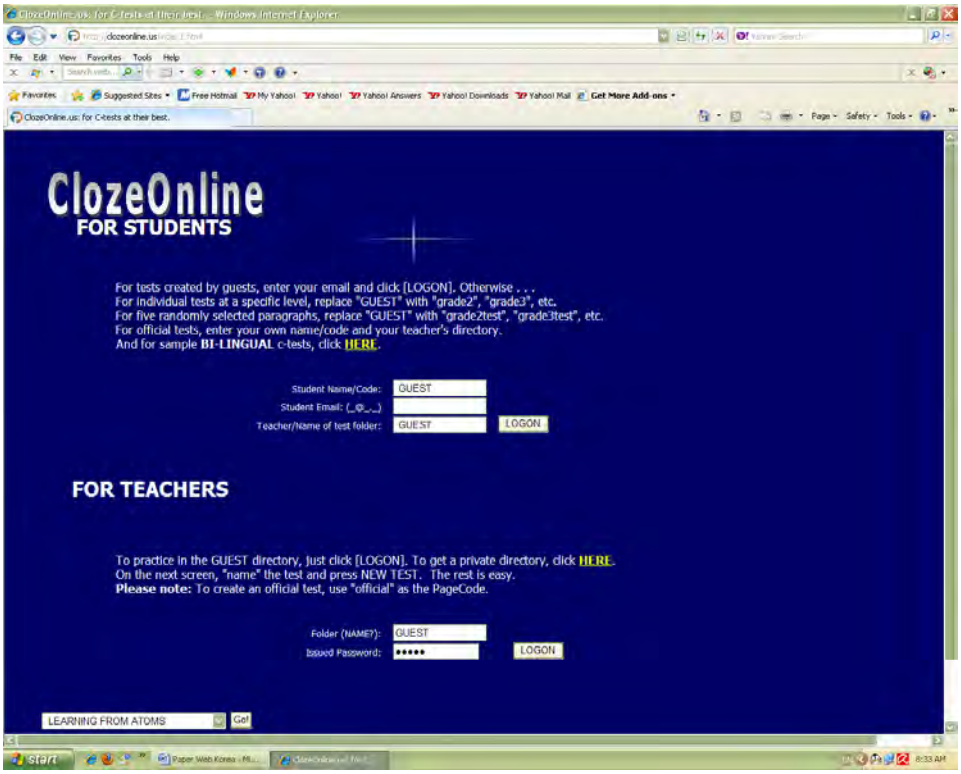
Boonsathorn, S. (2007). *C-Test revisited: A validation of Semantic/Syntactic Tests (S-Tests) for advanced students*. Frankfurt am Main: Peter Lang Frankfurt.

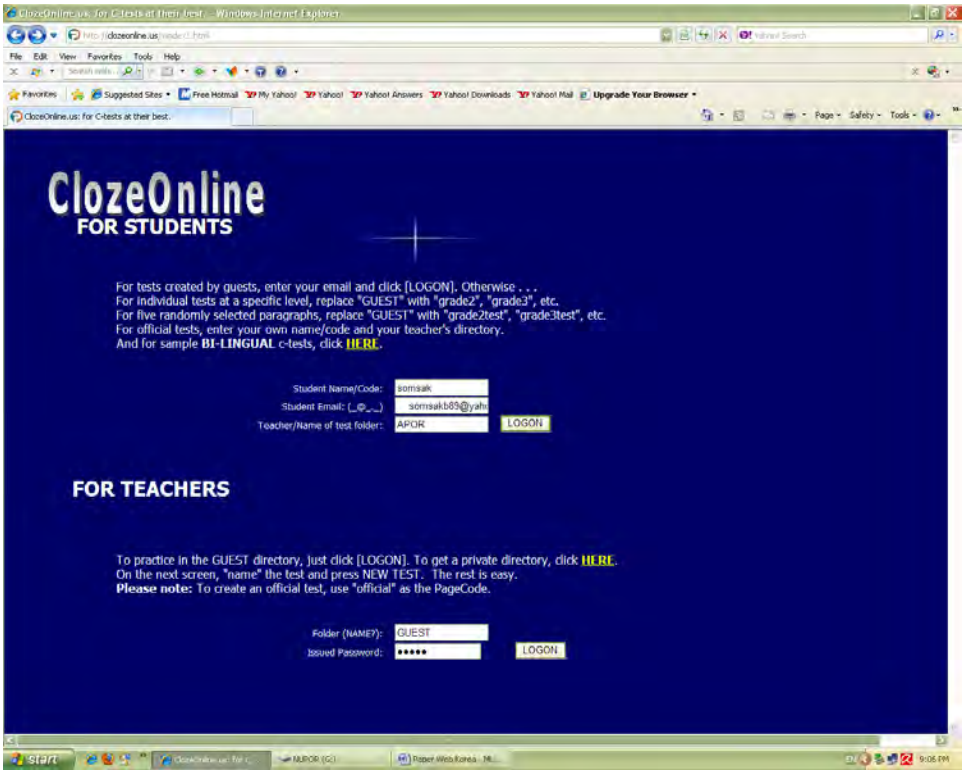
Boonsathorn, S., Getkham, K., & Boonsathorn, W. (2007). *Development of NIDA Test of English Competence (N-TEC)*. Research Report, The National Institute of Development Administration, Bangkok, Thailand.

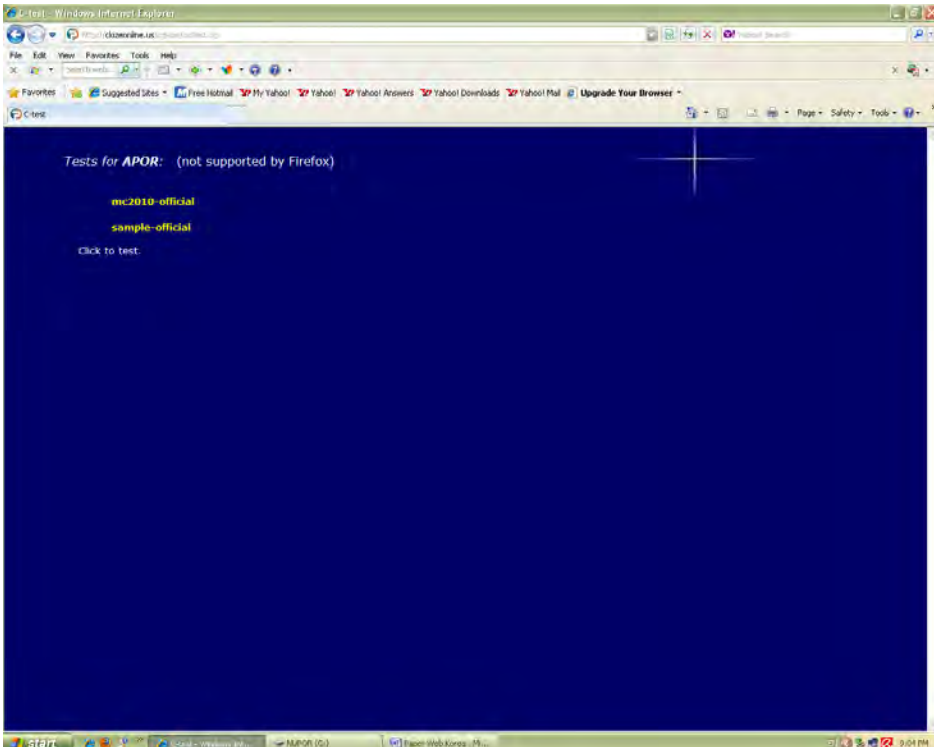
Bradshaw, J. (1990). Test-takers' reactions to a placement test. *Language Testing, 7*(1), 13-30. doi:10.1177/026553229000700103
Cleary, C. (1988). The C-Test in English: left-hand deletions. *RELC Journal, 19*(2), 26-35. doi:10.1177/003368828801900203
Cohen, A. D., Segal, M., & Bar-Siman-To, R. (1984). The C-Test in Hebrew. *Language Testing, 1*(2), 221-225. doi:10.1177/026553228400100206
Connelly, M. (1997). Using C-tests in English with post-graduate students. *English for Specific Purposes, 16*(2), 139-150. doi:http://dx.doi.org/10.1016/S0889-4906(96)00018-X
Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing, 9*(2), 187-206. doi:10.1177/026553229200900206
Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Literacy Research and Instruction, 6*(4), 126-135.
Kammasorn, N. (2008). *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 (PhD) --Chulalongkorn University, 2008.
Klein-Braley, C. (1981). Empirical investigations of cloze tests. *Unpublished PhD dissertation, University of Duisburg*.
Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: An appraisal. *Language Testing, 14*(1), 47-84.
Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-Test1. *Language Testing, 1*(2), 134-146. doi:10.1177/026553228400100202
Klein-Braley, C., & Stevenson, D. K. (1981). *Practice and problems in language testing*: Peter D. Lang.
Lazaraton, A., & Hatch, E. M. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.
Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement, 22*(4), 287-293.
Oller, J. W. (1976). Evidence for a general language proficiency factor: An expectancy grammar. *Die neueren sprachen, 75*(2), 165-174.
Oller, J. W. (1979). *Language tests at school: A pragmatic approach*: Addison-Wesley Longman Ltd.
Oller, J. W., & Conrad, C. A. (1971). The cloze technique and ESL proficiency. *Language Learning, 21*(2), 183-194.
Prapphal, K. (1994). A study of the C-Test and the X-Test performed by firstyear science-oriented university students. *PASAA, 24*, 16-23.
Prapphal, K. (1996). An investigation of the general and academic English X-Tests in measuring grammatical competence of Thai science students. *PASAA, 26*, 73-79.
Raatz, U., & Klein-Braley, C. (1981). The C-Test--A Modification of the Cloze Procedure.
Rouhani, M. (2008). Another look at the c-test: A validation study with Iranian EFL learners. *The Asian EFL Journal Quarterly 10* (1), 154-180.
Rungruangthum, M. (2005). *A comparative study between the C-Test and the NC-Test and between MC-Test and NMC-Test, using identical texts*. Mahidol University.
Sigott, G., & Kobrel, J. (1993). Validating the X-test. *Language Testing Update, 14*, 53-58.
Smith, F. (2012). *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read, Sixth Edition.* London: Taylor & Francis.
Spolsky, B. (1971). Reduced redundacy as a language testing tool. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics (Cambridge 8-2 September 1969)*. Cambridge: Cambridge University Press.
Stansfield, C., & Hansen, J. (1983). Field Dependence-Independence as a Variable in Second Language Cloze Test Performance*. *TESOL Quarterly, 17*(1), 29-38. doi:10.2307/3586422
Syndicate, U. o. C. L. E. (2001). *Quick Placement Test*: Oxford University Press.
Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journalism quarterly*.
Weir, C. J. (1990). *Communicative language testing*: Prentice Hall.
Wonghiransombat, P. (2013). The Cloze Test and the C-Test. *Thammasat Journal, 31*(2).

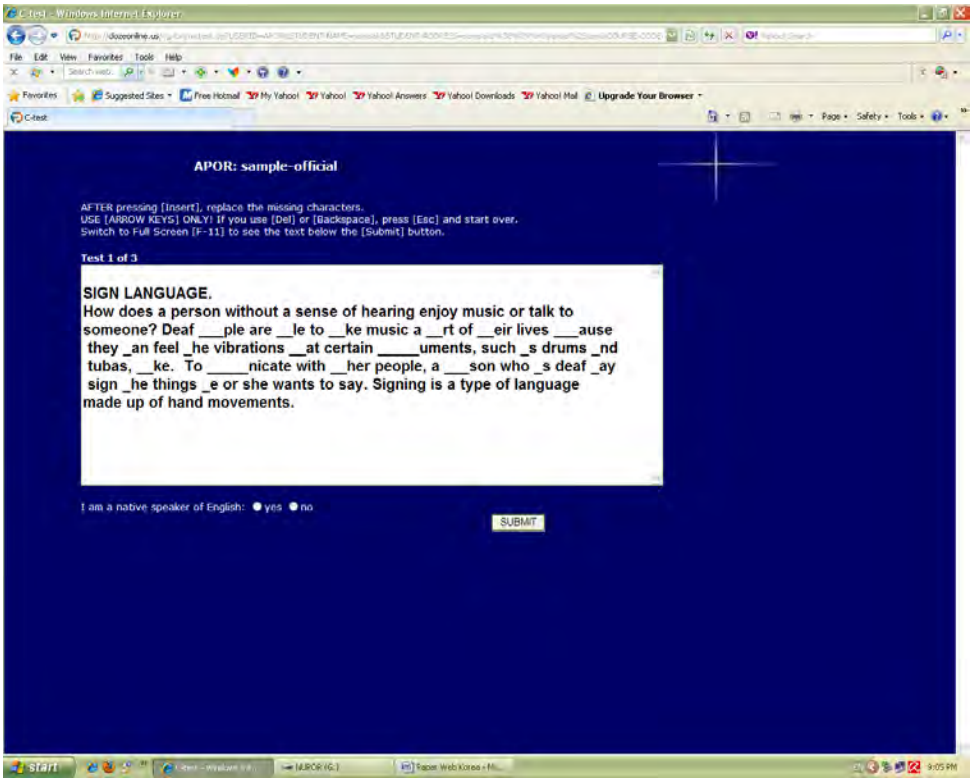
Appendix
Sample Test Access Screens

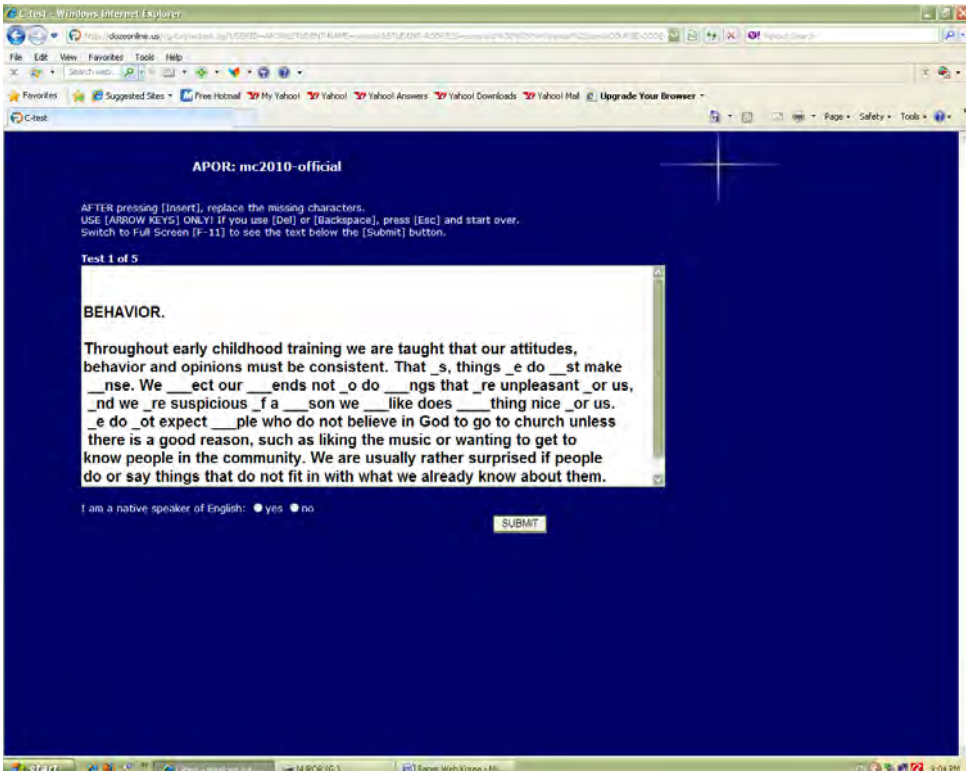





ClozeOnline
FOR STUDENTS
For tests created by guests, enter your email and click [LOGON]. Otherwise . . .
For individual tests at a specific level, replace "GUEST" with "grade2", "grade3", etc.
For five randomly selected paragraphs, replace "GUEST" with "grade2test", "grade3test", etc.
For official tests, enter your own name/code and your teacher's directory.
And for sample BI-LINGUAL c-tests, click HERE.
Student Name/Code:
Student Email: (_@_._)
Teacher/Name of test folder:
APOR
LOGON
FOR TEACHERS
To practice in the GUEST directory, just click [LOGON]. To get a private directory, click HERE.
On the next screen, "name" the test and press NEW TEST. The rest is easy.
Please note: To create an official test, use "official" as the PageCode.
Folder (NAME?):
GUEST
Issued Password:
LOGON


Tests for APOR: (not supported by Firefox)
mc2010-official
sample-official
Click to test.


APOR: sample-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.
Test 1 of 3
SIGN LANGUAGE.
How does a person without a sense of hearing enjoy music or talk to someone? Deaf ___ple are __le to __ke music a __rt of __eir lives ___ause they _an feel _he vibrations __at certain _____uments, such _s drums _nd tubas, __ke. To _____nicate with __her people, a ___son who _s deaf _ay sign _he things _e or she wants to say. Signing is a type of language made up of hand movements.
I am a native speaker of English: yes no
SUBMIT



Acknowledgements

The authors would like to thank Mae Fah Luang University for their financial support of the original project on the Web-based mC-Test and Dr. Arthur Dryver and Dr. Richard Lynch for their invaluable comments and feedback. They are also grateful to all anonymous reviewers and Dean Jorgensen for their insightful comments and suggestions.

About the Authors:

Somsak Boonsathorn received his PhD in TESL from The University of Alberta, Canada. He is now an assistant professor at Mae Fah Luang University, Thailand, in the Graduate Program of English for Professional Development. He initiated the mC-Test and developed the S-Test. Email: somsak.boo@mfu.ac.th.

Chayaporn Kaoropthai earned her M.A. (ELT) from King Mongkut's University of Technology Thonburi, Bangkok, Thailand, and is presently a PhD candidate in Educational Technology and Communications. Her research interests include innovations in education and language assessment. She is now with the School of Liberal Arts at Mae Fah Luang University. Email: Chayaporn.kao@mfu.ac.th.