

# *S-Tests: A Rational C-Test for Advanced Students\** **เอสเทสต์: ข้อสอบซีเทสต์แบบหลักเหตุผล** **สำหรับนักศึกษาระดับสูง**

Sumosak Boonsathorn, Ph.D.\*\*  
สมศักดิ์ บุญสาทร

## บทคัดย่อ

ข้อสอบเอสเทสต์ (S-Test) แตกต่างจากข้อสอบซีเทสต์ (C-Test) หรือข้อสอบเอ็มซีเทสต์ (MC-Test หรือ X-Test) ซึ่งใช้วิธีตัดคำออกครึ่งคำ โดยตัดคำเว้นคำและไม่คำนึงว่าคำเหล่านั้นเป็นคำชนิดใด S-Test ซึ่งย่อจาก Semantic/Syntactic Test เป็นข้อสอบ C-Test แบบหลักเหตุผล เนื่องจากใช้วิธีตัดคำออกครึ่งคำ โดยตัดเฉพาะคำบอกเนื้อหา (content word) คำเว้นคำ ผู้สอบจะต้องอ่านและเติมส่วนที่ตัดออกไปให้ถูกต้อง

S-Test ที่ใช้ในการศึกษาคำนี้แบ่งเป็นสองชนิด คือ S1-Test เป็นข้อสอบที่ตัดครึ่งแรกของคำ ส่วน S2-Test ตัดครึ่งหลังของคำ กลุ่มตัวอย่างที่ใช้ในการศึกษาคำนี้คือ นักศึกษาปริญญาโททางภาษาและการสื่อสารของสถาบันบัณฑิตพัฒนบริหารศาสตร์ ในภาคเรียนที่ 2 ปีการศึกษา 2541 จำนวน 97 คน โดยแบ่งออกเป็นสองกลุ่มที่มีความสามารถเท่าเทียมกันโดยใช้คะแนน TOEFL นักศึกษาเหล่านี้ได้ผ่านการคัดเลือกโดยใช้ข้อสอบคัดเลือกของสถาบัน

การวิเคราะห์ทางสถิติแสดงให้เห็นว่า S1-Test มีคุณสมบัติเหนือกว่า S2-Test ทั้งในด้านความเชื่อถือได้ (reliability) ความเที่ยงตรง (validity) และความเหมาะสมของระดับความยากง่าย (difficulty suitability) นอกจากนี้ ข้อมูลจากการตอบแบบสอบถามของผู้สอบยังสนับสนุนว่าข้อสอบ S-Test ทั้งสองชนิดมีความเที่ยงตรงประจักษ์ (face validity) อยู่ในระดับสูง

## Abstract

*Unlike in the "systematic" C-Test and MC-Test (a.k.a. X-Test), in which half of every second word (no matter what part of speech it is) is deleted, in the S-Test (Semantic/Syntactic Test-- a rational C-Test), only content words, i.e., every second "content" word is deleted.*

*The S1-Test, with the first half of the word deleted, and the S2-Test, with the second half of the word deleted were administered to two groups of graduate students in the Language and Communication program at NIDA in the second semester of 1998 (N=97). These students were divided according to their scores for a non-secure form of TOEFL so that their proficiency was controlled. They had also taken the NIDA Entrance Examination before they started their graduate program.*

*Statistical analyses revealed that the S1 Test was superior to the S2-Test with respect to reliability, concurrent validity, and difficulty suitability. The questionnaire data indicated that S-Tests have high face validity.*

\* This paper is a revised version of the original paper of the same title presented by the author at The Fourth International Conference on "EIT Collaboration: Towards Excellence in the New Millennium" at Chulalongkorn University, Thailand, December 1-3, 1999.

The author is grateful to the NIDA Research Promotion Endowment Fund for granting a research fellowship (April 1999 to March 2000) in support of the research project. He would also like to thank Ms. Patchara Wonghiransombot and Ms. Thitira Thongsa nga, his research assistants; Mr. Jeff Durand who made valuable suggestions; and all the students who participated and gave him their honest reactions.

\*\* Dr. Somsak Boonsathorn is presently a lecturer and co-director of the M.A. program in Language and Communication, the School of Language and Communication at the National Institute of Development Administration.

## INTRODUCTION AND BACKGROUND

S-Tests are the latest format of C-Test, proposed by the author to replace the original C-Test, which was initiated by Raatz & Klein-Braley (1981), as well as the MC-Test (a.k.a. X-Test), which was initiated by the author (Boonsathorn, 1987). The purpose of these tests is to measure general proficiency.

### The C-Test and the MC-Test

Based on the weaknesses found in the cloze test, Raatz & Klein-Braley (1981) proposed six criteria for a new test.

1. The new test should use several different texts.
2. It should have at least 100 deletions.
3. Educated native speakers should obtain virtually perfect scores.
4. The deletion 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.

Raatz & Klein-Braley used these six criteria to develop a new test, the C-Test. In the C-Test, four to six short passages (50-100 words each) are used. The deletion in each passage begins in the second sentence by deleting the second half of every second word and leaving the last sentence undeleted. The test taker's task is to read and complete the missing part of the word. The following is an example of the C-Test.

### C-Test (25 items)

*Tom was a student who wasn't interested in studying. He pref\_ \_ \_ \_ \_ to ha\_ \_ \_ a go\_ \_ \_ time. Natu\_ \_ \_ \_ \_ , when h\_ \_ took h\_ \_ \_ examinations h\_ \_ didn't g\_ \_ \_ good ma\_ \_ \_ . since h\_ \_ knew h\_ \_ \_ father wo\_ \_ \_ be an\_ \_ \_ with h\_ \_ , he se\_ \_ \_ a tele\_ \_ \_ \_ \_ to h\_ \_ \_ brother's ho\_ \_ \_ . In t\_ \_ \_ telegram h\_ \_ asked h\_ \_ \_ brother t\_ \_ \_ prepare h\_ \_ \_ father f\_ \_ \_ the b\_ \_ \_ news. The next morning, he recetived the following answer: "Father is prepared. You'd better prepare yourself."*

A modified C-Test, the MC-Test, was proposed by the author (Boonsathorn, 1987). The reason for proposing the MC-Test was that of the word formation of content words, i.e. nouns, verbs, adjectives and adverbs. Since most of them are in the form of BASE+ SUFFIX (ES), the first half and the second half of the word must be significantly different. While there is normally more of the semantic information in the first half, there is more of the syntactic information in the second half. Therefore, in the MC-Test, the author deleted the first half of every second word, instead, in order to examine semantic knowledge. The example using the same passage is as follows:

### MC-Test (25 items)

*Tom was a student who wasn't interested in studying. He \_ \_ \_ \_ \_ rrod to \_ ve a \_ \_ od time. \_ \_ \_ \_ \_ ally, when \_ e took \_ \_ s examinations \_ e didn't \_ \_ t good \_ \_ \_ ks. Since \_ knew \_ \_ s father \_ \_ \_ ld be*

...ry with ...m, he ...nt a ...gram  
 to ...s brother's ...se. In ...e telegram  
 ...e asked ...s brother ...o prepare ...s father  
 ...r the ...d news. The next morning, he  
 received the following answer: "Father is  
 prepared. You'd better prepare yourself."

The findings from the author's 1987 study, conducted in Alberta, Canada, with 389 high school students (L1) and 104 adult ESL students (L2) can be summarized as follows:

1. The C-Test and the MC-Test are highly reliable and valid for both L1 and L2 data.
2. The MC-Test is more difficult and discriminates better than the C-Test in both L1 and L2 data.
3. Different starting points of deletion do not affect the difficulty, reliability and validity of the tests.
4. Factor analysis results suggest that the C-Test and the MC-Test have different factor structures. The MC-Test seems to require more of the normal reading process than the C-Test does.
5. The Michigan Test correlates moderately highly with the C-Test/MC-Test. The C-Test, however, appears to predict performance on the Michigan Test better than the MC-Test does.
6. While the restoration of the C-Test passage relies primarily on the orthographic cues available in the items, interview results show that the processing of the MC-Test items necessitates the use of all the strategies required by readers in the process of normal reading.

### Why the S-Tests

The literature review reveals that many researchers (e.g. Raatz & Klein-Braley, 1981; Connelly, 1997; and Klein-Braley, 1997) find the C-Test a good, reliable, and valid test. On the other hand, those who find the C-Test too easy (e.g. Boonsathorn, 1987; Cleary, 1988; Sigott & Koberl, 1993; Prapphal, 1994, 1996; and Wonghiransombat, 1998) experimented with the MC-Test or X-Test and found that the MC-Test is more difficult, discriminates better, and therefore is more appropriate for advanced students.

Although most studies on the C-Test and the MC-Test appear to support that both tests are efficient, reliable, and valid, some weaknesses have been pointed out. Cleary (1988) and Dornyei & Kotona (1992) show that the "grammatically unmarked items" or the "function word items" are far too easy. Jafapur (1995) suggests that the difference in difficulty levels between content word items and function word items is considerable.

Since most function words are short, consisting of only one or two syllables, they are usually too easy in both the C-Test and the MC-Test, resulting in low discrimination power. The author, therefore, proposes to concentrate on only content words by deleting half of every second content word. And because most content words contain both semantic and syntactic information in themselves, the term S-Test (Semantic/Syntactic Test) will be used. The S-Test in which the first half of the word is deleted will be called the S1-Test, while the

one in which the second half is deleted will be called the S2-Test. The S-Tests are considered a rational C-Test because the author has a rationale in deleting only content words instead of systematically deleting every second word.

#### Statement of Purpose

The purpose of this study is to examine the S-Tests, a rational C-Test in which half of every second content word is deleted, focusing on:

1. The reliability of the S-Tests,
2. The difficulty suitability of the S-Tests,
3. The concurrent validity of the S-Tests, and
4. The face validity of the S-Tests.

## METHOD

### Subjects

The sample of this study was 97 graduate students in the Language and Communication part-time program, Class 1, in the second semester, 1998 at NIDA.

### Instruments

#### 1. S-Tests

Two parallel forms of S-Tests, the S1-Test with the first half of every second content word deleted and the S2-Test with the second half of every second content word deleted, were developed. Each form consisted of the same six short texts (110-172 words; readability, Flesch-Kincaid Grades 10.2 to 16.8), totaling 150 items. After piloting the tests, only

four texts (Grades 12 to 16.8) or 100 items were selected to be the final form of the tests. The examples of the S1 Test and the S2-Test constructed from the same text as the C-Test and the MC-Test are as follows:

#### S1-Test (12 items)

Tom was a student who wasn't interested in studying. He preferred to \_ \_ ve a good \_ \_ me. Naturally, when he \_ \_ ok his examinations, he didn't \_ \_ t good \_ \_ \_ ks. Since he knew his \_ \_ \_ her would be \_ \_ \_ ry with him, he sent a \_ \_ \_ \_ gram to his brother's \_ \_ \_ se. In the telegram he \_ \_ \_ ed his brother to \_ \_ \_ \_ are his father for the \_ \_ d news. The next morning, he received the following answer : "Father is prepared. You'd better prepare yourself."

#### S2-Test (12 items)

Tom was a student who wasn't interested in studying. He preferred to ha \_ \_ a good ti \_ \_ . Naturally, when he to \_ \_ his examinations, he didn't g \_ \_ good ma \_ \_ \_ . Since he knew his fat \_ \_ \_ \_ would be an \_ \_ \_ with him, he sent a tele \_ \_ \_ \_ to his brother's ho \_ \_ \_ . In the telegram be as \_ \_ \_ his brother to pre \_ \_ \_ \_ his father for the b \_ \_ news. The next morning, he received the following answer: "Father is prepared. You'd better prepare yourself."

#### 2. Proficiency Test

A non-secure form of TOEFL was used to measure students' general proficiency. There were three sections-Listening

Comprehension, Structure and Written Expression, and Reading Comprehension.

3. NIDA Entrance Examination for the Language and Communication Program

The test included four sections-Listening Comprehension, Error Identification, Cloze Test, and Reading Comprehension. In this study, the author used the students' Entrance Examination scores, which are readily available.

4. Questionnaire

A Likert-Scale nine-item questionnaire was developed by adapting Jafarpur's (1995) questionnaire and the author's (Boonsathorn, 1987) interview schedule. The purpose of the questionnaire was to establish the face validity of the S-Tests.

**Procedure**

All students who participated in the study took the non-secure form of TOEFL in the second semester of 1998. Their TOEFL scores were then used as a basis for dividing them into two equated groups in terms of proficiency level.

Two weeks later, Group 1 took the S1-Test while Group 2 took the S2-Test.

**Analyses**

The results of the S1-Test and the S2-Test and their corresponding TOEFL scores and Entrance Examination scores were separately analyzed by form using the Statistical Package for Social Science (SPSS/PC+6.1.4). The statistics obtained from these analyses were:

1. Cronbach's alpha,

2. Means and standard deviations,

3. Pearson's product-moment correlation coefficients, and

4. Percentage points and mean scores of questionnaire data.

**RESULTS**

Based on the purpose of the study, the results will be displayed in both tabular and descriptive form.

1. The Reliability of the S-Tests

**Table 1**

**The Reliability of S1-Test and S2-Test**

| Test form | <u>n</u> | Alpha |
|-----------|----------|-------|
| S1-Test   | 50       | .8740 |
| S2-Test   | 47       | .8043 |

From Table 1, the alpha of the S1-Test is 0.8740 while the alpha of the S2-Test is 0.8043. The S1-Test appears to be more reliable than the S2-Test; however, both can be considered reliable.

2. The Difficulty Suitability of the S-Tests

**Table 2**

**Means and t-test Statistics**

| Form    | <u>n</u> | <u>M</u> | <u>SD</u> | t      | p        |
|---------|----------|----------|-----------|--------|----------|
| S1-Test | 50       | 59.60    | 10.52     | -9.755 | .000     |
| S2-Test | 47       | 76.47    | 8.95      |        | (p<.001) |

Table 2 shows that the S1-Test is more difficult (M=59.60, S.D. = 10.52) than the S2-Test (M=76.47, S.D. = 8.95). Obviously, the difference in terms of difficulty is significant (p<.001).

Table 3

## Item Discrimination

| Discrimination Interval | S1-Test<br>(No. of items) | S2-Test<br>(No. of items) |
|-------------------------|---------------------------|---------------------------|
| > .75                   | 1                         | 5                         |
| .51 to .75              | 8                         | 9                         |
| .26 to .50              | 26                        | 17                        |
| < .26                   | 65                        | 69                        |
| Total                   | 100                       | 100                       |

In relation to the discrimination power of the tests, Table 3 indicates that the number of items whose discrimination indices are low (<.26) is a little smaller in the S1-Test (65) than in the S2-Test (69). However, the number of items with very high discrimination indices (>.75) is greater in the S2-Test (5) than in the S1-Test (1).

### 3. The Concurrent Validity of the S-Tests

Table 4

## Correlation Coefficients between S1-Test Scores, the TOEFL Scores and the NIDA Entrance Examination (ENTR) Scores

|         | S1-Test  | TOEFL  | ENTR   |
|---------|----------|--------|--------|
|         | (n = 50) |        |        |
| S1 Test |          | .6211* | .5762* |
| TOEFL   |          | --     | .6894* |
| ENTR    |          |        | --     |

\*p = .000 (p<.001)

Table 5

## Correlation Coefficients between S2-Test Scores, the TOEFL Scores and the NIDA Entrance Examination (ENTR) Scores

|         | S2-Test | TOEFL  | ENTR   |
|---------|---------|--------|--------|
|         | (n=47)  |        |        |
| S2-Test | --      | .6057* | .5619* |
| TOEFL   |         | --     | .6462* |
| ENTR    |         |        | --     |

\*p = 0.000 (p<.001)

the information from Tables 4 and 5 reveals that the scores of the S1-Test and the S2-Test both fairly highly correlate with the TOEFL scores (0.6211 and 0.6057) and the NIDA Entrance Examination scores (0.5762 and 0.5619). The S1-Test scores appear to correlate better with both the TOEFL scores and the NIDA Entrance Examination scores than the S2-Test scores do.

The correlation coefficients between the TOEFL scores and the NIDA Entrance Examination scores are moderately high in both the S1-Test subjects (0.6894) and the S2-Test subjects (0.6462).

### 4. The Face Validity of the S-Tests

**Table 6**

**Questionnaire Data of the S1-Test in Percentage Points and Mean Scores**

(1 = strongly disagree; 5 = strongly agree).

|  | 1     | 2     | 3     | 4     | 5     | M    |
|--|-------|-------|-------|-------|-------|------|
| 1. The S1-Test is a language test.   | 2.00  | 4.00  | 18.00 | 28.00 | 48.00 | 4.16 |
| 2. It is an interesting test.  | 2.04  | 6.12  | 8.17  | 53.06 | 30.61 | 4.04 |
| 3. It is a challenging test.   | 0     | 6.12  | 2.04  | 42.86 | 48.98 | 4.35 |
| 4. The S1-Test measures:   |       |       |       |       |       |      |
| 4.1 Vocabulary.  | 2.04  | 0     | 2.04  | 30.61 | 65.31 | 4.57 |
| 4.2 Grammar.   | 6.12  | 12.24 | 38.78 | 22.45 | 20.41 | 3.39 |
| 4.3 Reading.   | 4.08  | 6.12  | 12.25 | 40.82 | 36.73 | 3.77 |
| 4.4 General English Proficiency.   | 2.04  | 0     | 14.29 | 53.06 | 30.61 | 4.10 |
| 4.5 Background knowledge.  | 0     | 4.08  | 2.04  | 30.61 | 63.27 | 4.53 |
| 4.6 Intelligence.  | 4.17  | 6.25  | 14.58 | 43.75 | 31.25 | 3.92 |
| 4.7 Other (specify): _____   |       |       |       |       |       |      |
| 5. It is a fair test of English.   | 6.12  | 16.33 | 40.82 | 26.53 | 10.20 | 3.18 |
| 6. It is a good test of English.   | 4.08  | 12.24 | 36.75 | 36.73 | 10.20 | 3.37 |
| 7. It is fun to do this type of test.  | 8.16  | 2.04  | 10.21 | 42.86 | 36.73 | 3.98 |
| 8. It causes a lot of anxiety.   | 14.29 | 16.33 | 16.31 | 38.78 | 14.29 | 3.22 |
| 9. The S1-Test can be used as:   |       |       |       |       |       |      |
| 9.1 A general English Proficiency test.  | 10.20 | 10.20 | 30.62 | 32.65 | 16.33 | 3.35 |
| 9.2 Part of the entrance examination for the Language and Communication program. | 12.24 | 12.24 | 62.67 | 30.61 | 30.61 | 3.18 |

**The S1-Test**

The questionnaire data in Table 6 suggest that most respondents agree to strongly agree that the S1-Test is a language test (M=4.16) which is interesting (M=4.04) and challenging (M=4.35). They believe that the S1-Test mainly measures vocabulary (M=4.57), background knowledge (m=4.53), and general English proficiency (M=4.10). They rather agree

that it is fun to do the S1-Test (M=3.98), while they are not quite certain whether it is a good test of English (M=3.37), a fair test of English (M=3.18), and it causes a lot of anxiety (M=3.22). The respondents, however, do not quite agree that the S1-Test can be used as a general English proficiency test (M=3.35) or part of the entrance examination for the Language and Communication program (M=3.18).

Table 7

Questionnaire Data of S2-Test in Percentage Points and Mean Scores

(1 = strongly disagree; 5 = strongly agree).

|  | 1     | 2     | 3     | 4     | 5     | <u>M</u> |
|--|-------|-------|-------|-------|-------|----------|
| 1. The S2-Test is a language test.   | 4.26  | 8.15  | 12.76 | 14.89 | 59.57 | 4.17     |
| 2. It is an interesting test.  | 6.25  | 6.25  | 25.00 | 25.00 | 37.50 | 3.81     |
| 3. It is a challenging test.   | 6.38  | 8.51  | 17.03 | 23.40 | 44.68 | 3.91     |
| 4. The S2-Test measures:   |       |       |       |       |       |          |
| 4.1 Vocabulary.  | 4.26  | 2.13  | 10.63 | 12.77 | 70.21 | 4.43     |
| 4.2 Grammar.   | 4.17  | 4.17  | 20.83 | 19.17 | 41.67 | 4.00     |
| 4.3 Reading.   | 4.17  | 2.08  | 20.83 | 22.92 | 50.00 | 4.13     |
| 4.4 General English Proficiency.   | 2.17  | 4.35  | 23.91 | 19.57 | 50.00 | 4.11     |
| 4.5 Background knowledge.  | 4.26  | 4.26  | 21.27 | 27.66 | 42.55 | 4.00     |
| 4.6 Intelligence.  | 6.52  | 6.52  | 28.27 | 23.91 | 34.78 | 3.74     |
| 4.7 Other (specify): _____   |       |       |       |       |       |          |
| 5. It is a fair test of English.   | 8.33  | 12.50 | 35.42 | 35.42 | 8.33  | 3.23     |
| 6. It is a good test of English.   | 8.33  | 18.75 | 31.25 | 18.75 | 22.92 | 3.29     |
| 7. It is fun to do this type of test.  | 6.25  | 8.33  | 12.51 | 33.33 | 39.58 | 3.92     |
| 8. It causes a lot of anxiety.   | 19.57 | 26.09 | 34.78 | 6.52  | 13.04 | 2.67     |
| 9. The S2-Test can be used as:   |       |       |       |       |       |          |
| 9.1 A general English Proficiency test.  | 6.25  | 14.83 | 16.66 | 41.67 | 20.83 | 3.56     |
| 9.2 Part of the entrance examination for the Language and Communication program. | 10.42 | 20.83 | 20.84 | 27.08 | 20.83 | 3.27     |

## The S2-Test

Table 7 shows that in general the respondents agree that the S2 Test is a language test (M=4.17) which is rather interesting (M=3.81) and challenging (M=3.91). They consider that the S2-Test mainly measures vocabulary (M=4.43), reading (M=4.13), and general English proficiency (M=4.11). They rather agree that it is fun to do

the S2-Test (M=3.92), but they are not sure whether it is a good test (m=3.29) and a fair test (M=3.23) of English. The S2-Test does not seem to cause as much anxiety (M=2.67) as the S1-Test. The respondents tend to agree that the S2-Test can be used as a general proficiency test (M=3.56) rather than part of the entrance examination for the Language and Communication program (M=3.27).



## DISCUSSION AND CONCLUSION

Statistical analyses reveal that both the S1-Test and the S2-Test have high reliability, which is one of the most important qualities of tests. The S1-Test is more reliable than the S2-Test. The finding supports the previous studies (e.g., Boonsathorn, 1987; Cleary, 1988) that the MC-Test is more reliable than the C-Test.

The S1-Test is more difficult than the S2-Test as expected. The S1-Test, with a mean of 59.60, seems to be more appropriate for these graduate students than the S2 Test, since the mean of 76.47 shows that it is probably too easy. Thus, the texts with Flesch-Kincaid Grades of 12 to 17 are probably appropriate for Thai graduate students. Item discrimination indices also support that the S1-Test has generally higher discrimination power than the S2-Test.

Correlational analyses indicate that both the S1-Test scores and the S2-Test scores can predict the performance on the TOEFL and the NIDA Entrance Examination to a certain extent, in spite of the different natures of the tests. The S1-Test scores correlate better than the S2-Test scores. This is contradictory to the author's 1987 study. The correlation coefficients between the TOEFL scores and the NIDA Entrance Examination scores are rather high, probably because they have a similar format.

The information obtained from the questionnaire shows that the respondents generally agree that both the S1-Test and the S2-Test are a language test which is interesting and challenging. The S1 Test is considered more challenging than the S2-Test. The respon-

dents agree that the S1-Test mainly measures vocabulary, background knowledge and general English proficiency while the S2-Test mainly measures vocabulary, reading, and general English proficiency. They rather agree that it is fun to do this type of test but are not quite sure that it is a good and fair test of English or not. The S1-Test causes more anxiety than the S2-Test does. The respondents are not sure whether the S1-Test and the S2-Test should be used as a general English proficiency test or part of the entrance examination or not. In general, the S1-Test and the S2-Test appear to have high face validity. The findings are completely contradictory to Jafarpur's (1995) results which show that the C-Test has no face validity.

In conclusion, both the S1-Test and the S2-Test, have been supported to be reliable and valid tests. The S1-Test, however, appears to be superior to the S2-Test in all respects—reliability, concurrent validity, face validity, and difficulty suitability.

In order to verify whether different starting points of deletion of every second content word affects the reliability, validity and difficulty level of the test, the author is now further experimenting with the S1-Test with a new group of graduate students. Two parallel forms of the S1-Test—Form A and Form B—will be compared. Form A is the original form of the S1-Test where deletion starts from the second content word of the second sentence, while in Form B deletion starts from the third content word. The results will help confirm whether deletion in the S1-Test affects a representative sample of text or not.

## REFERENCES

- Boonsathorn, S. (1987). C-Test, proficiency, and reading strategies in ESL. Unpublished doctoral dissertation. University of Alberta, Alberta, Canada.
- Cleary, C. (1988). The C-Test in English : Left-hand deletion. *RELC Journal*, 19(2), 26-38.
- Connelly, M. (1997). Using C-Test in English with post-graduate students. *English for Specific Purposes*, 16(2), 139-150.
- Domyei, Z. & Katona, L. (1992). Validation of the C-Test amongst Hungarian EFL learners. *Language Testing*, 9(2), 187-206.
- Jafarpur, A. (1995). Is C-testing superior to cloze? *Language Testing*, 10(1), 195-216.
- Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: An appraisal. *Language Testing*, 14(1), 47-84.
- 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. (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. In T. Culhane, C. Klein-Braley, & D.K. Stevenson (Eds.), *Practice and problems in language testing*. Department of Language and Linguistics Occasional Papers No. 26. Colchester: University of Essex, pp. 113-138.
- Sigott, G. & Koberl, J. (1993). Validating the X-Test. *Language Testing Update*, 14, 53-58.
- Wonghiransombat, P. (1998). A validation of the MC-Test as a test of English proficiency. Unpublished master's research paper. The National Institute of Development Administration, Bangkok, Thailand.