

Articles

Testing English Tests: A Language Proficiency Perspective

Akihiro Ito

Hiroshima University

This study examines whether English questions on Japanese university entrance examinations are reliable and valid measures of examinees' language proficiency. The following tests were administered to 100 college freshmen: a narration-based, 700-word level, 70-item, open-ended Cloze Test; a test from the National Center for University Entrance Examination (NCUEE-Test) with the additional paper-pencil Pronunciation Tests; three grammar tests (A, B, C); TOEFL Listening Comprehension Test. Results indicate the NCUEE-Test is a fairly reliable and somewhat valid testing device to measure students' English proficiency. The grammar tests had a tendency not to be reliable and valid measures of students' proficiency, though the Grammar Test C showed moderate reliability and correlation with the Cloze Test. The Pronunciation Test in the NCUEE-Test is not a reliable measure of the students' listening ability. The need to develop a reliable external criterion listening test and to replicate the same kind of research using a larger sample of participants is also discussed.

本研究は、日本の大学入試の英語の問題が受験者の言語能力を測定する信頼性と妥当性をもっているかどうかを検証するものである。以下のテストを、100名の大学1年生に実施した。700語レベルの物語を使った70項目の書き込み式クローズテスト、大学入試センター試験に文字による発音問題を追加したもの、3種類の文法問題（A、B、C）、TOEFLのリスニング問題である。その結果、大学入試センター試験はかなり信頼性があり、学生の英語能力を測定する手段として妥当性がないわけではないこと、文法のテストは、Cのみならず信頼性とクローズ・テストとの相関を示したものの、信頼性、妥当性とも低い傾向にあること、大学入試センター試験の発音問題は、学生の聞き取り能力を測る手段としては信頼性がないことがわかった。最後に、信頼性のある外部の基準に基づいた聞き取りのテスト開発と、より多数の被験者を使った本研究の追試の必要性が論じられている。

English tests are widely used for screening students in university entrance examinations in Japan. However, since some regard the tests as immutable and limited empirical research on the reliability and validity of English questions has been conducted, the conclusions from the research tend to be speculative. I would therefore like to shed some light on the reliability and validity of English questions used for university entrance examinations, and to provide data for further improvement in the quality of the questions.

The types of English entrance examinations may be divided into the following three categories: (a) the test of the National Center for University Entrance Examination (NCUEE-Test), (b) the second screening by the national university entrance examinations, and (c) private university entrance examinations.

Problems with English Tests in University Entrance Examinations

In this section, the problems with the NCUEE-Test, the second screening tests for national university entrance examinations, and the private university entrance examinations will be clarified and the Group of Hypotheses will be set up for the investigation.

Concerns about the NCUEE-Test: In order to enter a national university in Japan, examinees must take two examinations: the NCUEE-Test and the second screening test at their prospective university.

In 1979, the *Kyōtsū Ichiji* (Preliminary Standard College Entrance Examination: PSCEE) was introduced by the Japanese Ministry of Education, Science, and Culture to Japanese universities as an entrance examination. The purpose of the PSCEE was to measure applicants' basic achievement levels in various subjects. In 1989, the university examination system changed and the PSCEE was altered to a different type of test: the NCUEE-Test. However, the contents remained virtually the same as those of the PSCEE. The NCUEE-Test is a "a first stage exam, somewhat analogous to the College Board SAT, in that many universities subscribe to it" (Brown & Yamashita, 1995, p. 12).

There seem to be fewer problems with the reliability and validity of English tests in the PSCEE and the NCUEE-Test in comparison with other English tests. Historically speaking, several researchers have proposed improvement in the validity and reliability of the NCUEE-Test. Kiyomura (1989, p. 245) argues, as a result of his inquiry, that the English questions in the NCUEE-Test are relatively valid for measuring students' achievement level. He found a moderate correlation ($r = 0.330$ to 0.620) between the

examinees' scores in the NCUEE-Test and those scores appearing in the records from their high schools. In addition, Yanai, Maekawa, and Ikeda (1989) report constantly high reliability coefficients of English tests ($r = 0.940$ to 0.956) in the PSCEE from 1979 to 1984.

However, there is one very serious problem with the test. Although fifteen years have passed since the PSCEE/NCUEE-Test was first used, a listening test has not been administered. The lack of the listening test and the problems with the use of paper-pencil pronunciation tests as alternative measures have been widely and heatedly discussed (Ishii, 1981; Kira, 1981; Kuniyoshi, 1981; Masukawa, 1981; H. Suzuki, 1981; Kashima, Tanaka, Tanabe, & Nakamura, 1983; Ohtomo, 1983; Shiozawa, 1983; Ibe, 1983; S. Suzuki, 1985; Ikeura, 1990; T. Takahashi, 1990; Wakabayashi & Negishi, 1990a, 1990b, 1991, 1994).

It seems that most researchers and university personnel are concerned with three issues. First, that the pronunciation test in the first section of the PSCEE or NCUEE-Test is not a reliable or valid measure of examinees' listening and pronunciation abilities. Second, that though it is very expensive to acquire the supplies needed to conduct a listening test nation-wide, a listening test is a "must" to produce a more valid PSCEE (NCUEE-Test) (H. Suzuki, 1981, pp. 23-25). Third, even though the Ministry of Education, Science, and Culture (1989) implemented aural/oral guidelines in 1994 and has tried to enhance high school students' listening skills in Oral Communication B, the English tests in the NCUEE-Test have not changed to reflect the new direction toward emphasizing aural skills. In short, as Brown and Yamashita (1995, p. 28) say, there is a contradiction between what is tested in the NCUEE-Test and what the Ministry of Education promotes in its curriculum.

It is hoped, therefore, that the NCUEE-Test will institute the use of a listening component as soon as possible to increase the validity of the tests, just as the STEP (The Society for Testing English Proficiency) test has already done (Shimizu, 1989, p.115).

In sum, a debate has been entertained regarding the lack of listening components in the NCUEE-Test, but the issue has remained unresolved because of various factors, including expense.

Concerns about the second screening tests for national university entrance examinations: In contrast to the NCUEE-Test, the national university second screening tests are still of the translation type, though some researchers (Wakabayashi & Negishi, 1994) argue that translation cannot be a reliable and valid measure of examinees' English proficiency. Brown and Yamashita (1995) point out that translation items in university entrance

examinations are "out of date because translation was abandoned years ago in ESL instruction" (p. 28). In addition, the test called Sogo Mondai (General Questions) has been criticized by Utsunomiya (1985) and Wakabayashi and Negishi (1990a, 1990b, 1991, 1994). For this paper, however, research on the second screening tests for national university entrance examinations was not be conducted. This is because it is quite difficult to get information about the scoring criteria for these tests from each university. This research paper targets the objective tests, such as the NCUEE-Test and grammar tests used in private university entrance examinations.

Concerns about the private university entrance examinations: The concerns about the English questions on private university entrance examinations are divided into the following two categories. First, in some examinations there are several types of questions, such as grammar tests, which seem to measure examinees' knowledge about intricate grammatical rules (Wakabayashi & Negishi, 1990a, 1990b, 1991, 1994) not language proficiency as a whole. Second, a large number of questions are too difficult for the examinees, who are mainly third-year high school students (Negishi, 1990).

The hypotheses: From the above, the following three Groups of Hypotheses (GH) were developed.

- GH 1: The reliability of the NCUEE-Test will be high. There will be a moderate correlation between the external criterion test, the Cloze Test, and the NCUEE-Test.
- GH 2: The reliability of the Pronunciation Test will be low. There will be correlation, albeit low, between the external criterion listening test, the TOEFL Listening Test, and the Pronunciation Test.
- GH 3: The reliability of grammar tests used in private university entrance examinations will be low. There will be correlation, again low, between the scores of grammar tests and that of the Cloze Test.

The Study

As the purpose of this study was to investigate the reliability and validity of English questions on Japanese university entrance examinations, my goals were: 1) to determine if the English questions in the Japanese university entrance examinations, such as the NCUEE-Test, the paper-pencil Pronunciation Test, and the grammar tests, are reliable and

valid measures of students' language proficiency, and 2) to examine the three groups of hypotheses set up in the introduction.

Method

Subjects: The experimental sample ($n = 100$) was taken from the first-year students who were enrolled in an undergraduate class in general English at Aichi University of Education in Japan. Most of them were eighteen years old. They were predominantly male. All of them would have taken more than six years of formal English courses prior to this study. They were majoring in a scientific field. The sample was thus homogeneous with regard to nationality, language background, educational level, and age. It is noteworthy that the participants were cooperative and showed a great deal of interest in the research.

Instruments: The following instruments were used in this experiment:

1. A 70-item open-ended Cloze Test (Appendix). The participants were allowed 30 minutes for completion.
2. A 50-item TOEFL Listening Comprehension Test (Steinberg, 1987, pp. 75-89). It took 25 minutes to finish this listening test.
3. English questions used in university entrance examinations.
 - a. A 58-item NCUEE-Test; 1991, the second version.
 - b. A 20-item Pronunciation Test; 1989 & 1992. The Pronunciation Test was administered with the NCUEE-Test. The Pronunciation Tests in NCUEE-Tests are basically divided into three types. In this study, the paper-pencil pronunciation test where participants are required to distinguish the segmental phonemes was used. Sixty minutes were allowed to complete 3a & 3b.
 - c. 16-item grammar tests A, B, and C: 1991, 1992, & 1993. The participants were allowed 25 minutes for completion.

Grammar Test categorization: The types of the grammar tests used in this study are divided into two categories, Grammar Test A type and Grammar Test C type, following Wakabayashi & Negishi (1990a, 1990b, 1991). Grammar Test A, called *Goi Hoju Mondai*, is a fill-in-the-blank-with-an-appropriate word composition test, whereas Grammar C, *Seijo Mondai*, is a placing-the-words-in-order composition test. The Grammar Test B consists of two parts, both fill-in-the-blank and placing-the-words-in-order.

Examples of the Grammar Tests: In Grammar Test A examinees are required to fill in the blanks by selecting one of four given words.

Example A: The two players faced each other () the chessboard in the final match.

- (A) along (B) to (C) toward (D) across

In Grammar Test B the examinees are asked to fill in the blanks by selecting one of four given words in first half.

Example B1: They could not _____ the expense of sending their daughter to college.

- A. give B. afford C. spoil D. spend

In the second half, the examinees are required to place given words in their correct order and to mark the number of words which are put in the second and fourth blanks, following a Japanese translation as a key.

Example B2: Kono shosetsu no sakusha wa daredato omoimasuka (in Japanese).

Who () (1) () (2) () () of the novel?

- [A. is B. do C. the D. you E. think F. author]
 (1. CA 2. CF 3. DA 4. DC 5. EA)

In Grammar Test C, first the examinees are required to place given words in the correct order to produce meaningful sentences, using a Japanese translation as a key. However, one of the words essential to make a meaningful sentence is intentionally eliminated. Then the examinees must select an appropriate word from the given word list to complete the sentence.

Example C: Karewa nisan nichi de shinu desho (in Japanese).

(he, few, will, die, days, a)

- A. maybe B. after C. in D. later

The NCUEE-Test, the additional Pronunciation Tests, Grammar Tests A, B, and C, and their answers were sampled from the *Daigaku Nyushi* (University Entrance Exams) Series (1994). The grammar tests were selected from three different university tests published in the series.

Procedure

Cloze Test construction & scoring method: In order to examine the concurrent validity of the NCUEE-Test and grammar tests in question, a carefully constructed Cloze Test was used as an external criterion test. The Cloze Test was produced on the basis of the results of recent research on cloze test construction: 1) the selection of appropriate texts (Nishida, 1986, 1987; Mochizuki, 1984, 1994; Y. Takahashi, 1984, 1988); 2) word-level (Mochizuki, 1992); 3) the essential number of questions and scoring methods (Sciarone & Schoorl, 1989); and 4) the deletion frequency of words (Alderson, 1979; Nishida, 1985). The cloze passage was adapted

from a low intermediate reader (700 word level) for Japanese high school students by Ishiguro & Tucker (1989). The passage selected, "Wang's story," a relatively neutral, narrative topic, contained 457 words. Its readability level was about 8th grade level as measured by the Flesch-Kincaid readability formula by using computer program *Grammatik IV* (1988). The Cloze Test itself was created by deleting every 6th word for a total of 70 blanks. Two sentences were left intact: one was at the beginning of the passage and one at the end to provide complete context. Sciarone & Schooli (1989) say that a cloze test of about 75 items should be scored with the contextually acceptable method to maintain a satisfactory reliability ($r > .80$). Then the Cloze Test was scored by me based on the contextually acceptable word method with the help of a British native speaker of English.

Reliability and concurrent validity of the Cloze Test: In order to examine the concurrent validity of the Cloze Test itself, the correlation between the scores of the 100-item TOEFL Practice Test (40 items structure & written expression, 60 items reading comprehension) (Steinberg, 1987), and the Cloze Test was measured in a pilot study. In the pilot study, the participants were 100 sophomore students enrolled in general English at Aichi University of Education in Japan. They were all Japanese speakers and majoring in Japanese language education or art education. Most of them were eighteen years old. The ratio between male and female was almost 1 to 1.

In the pilot study and the main study the split-half method was used to calculate the reliability of the tests. The use of the split-half method in the calculation of the reliability of cloze tests is still controversial because the method basically is designed for estimating the reliability coefficient of tests in which each item is independent. Brown (1983), however, shows that the lack of independence between items was not a problem for the internal consistency estimates on cloze tests. So I felt that the split-half method was permissible for estimating the reliability of cloze tests. In the split-half method, I scored the odd and even numbered items separately and examined the correlation between two halves. After the value for r was measured, each value for r was corrected for the reduction to half-test length using the Spearman-Brown prophecy formula ($r_{xx'} = 2r_{hh}/(1+r_{hh})$).

Results of the Pilot Study

High reliability of the tests and the moderate correlation coefficient between them ($r = 0.489$, $p < 0.01$) and the correlation coefficient cor-

Table 1: Reliability Coefficients by Spearman-Brown Split-Half Method ($n = 100$)

Tests	r	Mean(M)	Full Score	SD
Cloze Test	0.840	33.450	70	7.815
TOEFL	0.781	31.720	100	8.128

rected for attenuation in the Cloze Test is: $r = 0.604$, $p < 0.01$ were found. In this regard, the Cloze Test had a relatively high reliability coefficient ($r = 0.840$) and moderate correlation ($r = 0.489$) with a reliable discrete-point test such as TOEFL¹ (Table 1).

The final decision on the Cloze Test: The author then made a final decision to employ the Cloze Test as an external criterion test for measuring participants' language proficiency in this investigation. This is because the reliability of the Cloze Test exceeded the critical threshold level of 0.80 ($r = 0.840$) and the test correlated with the reliable discrete-point test, TOEFL, at close to 0.5 ($r = 0.489$, $p < 0.01$). The correlation corrected for attenuation in the Cloze Test was more than 0.5 ($r = 0.604$, $p < 0.01$).¹

The reliability and concurrent validity of the TOEFL Listening Test: The TOEFL Listening Comprehension Test was used as another external criterion test for measuring the concurrent validity of the Pronunciation Test. The author did not conduct a study on the concurrent validity of the TOEFL Listening Test. The reason for this is that the test has been utilized in real TOEFL testing sessions, and we can therefore conclude that the test might be reliable and valid enough to be an external criterion listening test.

Scoring procedure for the other tests: After all the tests except the Cloze Test were administered, the test papers were exchanged between students and scored under my direction. After the test papers were collected, they were reviewed by me before the statistical calculations were performed.

Results of Study

Table 2 shows the reliability coefficients of the tests. The reliability coefficients of the Cloze Test ($r = 0.853$), slightly higher than in the pilot study, and the NCUEE-Test ($r = 0.817$) are high. The three grammar tests are placed in the order of reliability coefficients from highest to lowest:

Table 2: Reliability Coefficients by Spearman-Brown Split-Half Method ($n=100$)

Tests	r	Mean(M)	Full Score	SD
Cloze Test	0.853	32.850	70	7.697
NCUEE-Test	0.817	34.920	58	8.013
Grammar Test A	0.099	9.190	16	1.948
Grammar Test B	0.436	7.930	16	2.483
Grammar Test C	0.570	7.810	16	2.497
TOEFL				
Listening Test	0.398	13.060	50	3.484
Pronunciation Test	0.208	8.350	20	2.355

C ($r = 0.570$); B ($r = 0.436$); A ($r = 0.099$). The differences in mean scores among the three grammar tests show that the mean score of the Grammar Test A is the highest ($M = 9.190$), and the mean scores of the other two Grammar Test B and C are almost the same ($M = 7.930$; $M = 7.810$). Though it is clear that the participants performed better in Grammar Test A than in grammar tests B and C in terms of the mean scores, the reliability of Grammar Test A is the lowest among the three. The reliability coefficients of the TOEFL Listening Comprehension Test ($r = 0.398$) and the Pronunciation Test ($r = 0.208$) are low.

Table 3 displays the correlation coefficients between the Cloze Test and NCUEE-Test and grammar tests. There were moderate correlations between the Cloze Test and the NCUEE-Test and Grammar Test C and low correlations between the Cloze Test and grammar tests A and B. The three grammar tests are placed in the order of the correlation coefficients from highest to lowest: C ($r = 0.441$); A ($r = 0.346$); B ($r = 0.323$). Grammar Test C shows the highest reliability and correlation among the

Table 3: Correlation between Cloze Test and the NCUEE-Test and the Grammar Tests ($n=100$)

Tests	r	p
Cloze Test and NCUEE-Test	0.462	<0.01
Cloze Test and Grammar Test A	0.346	<0.01
Cloze Test and Grammar Test B	0.323	<0.01
Cloze Test and Grammar Test C	0.441	<0.01

grammar tests. Finally, no significant correlation was found between the TOEFL Listening Comprehension Test and the Pronunciation Test in this study ($r = -.078$, $p = n.s.$).

Discussion

The results were then examined according to the three Groups of Hypotheses (GH) presented above.

GH 1, of a high reliability for the NCUEE-Test ($r = 0.817$) and moderate correlation ($r = 0.462$, $p < 0.01$) between the Cloze Test and NCUEE-Test, was supported. Therefore, the NCUEE-Test is a fairly reliable and somewhat valid measure of examinees' English proficiency.

GH 2, of low reliability for the Pronunciation Test ($r = 0.208$), was supported. However, that part of GH 2 which posited correlation, albeit low, between TOEFL Listening Test and the Pronunciation Test was not supported because there was negative correlation, although not significant, between the two tests ($r = -0.078$, $n.s.$). The lack of correlation may be due to the fact that the ability to distinguish among segmental phonemes in the Pronunciation Test cannot cover the wide range of listening abilities which the TOEFL Listening Test tries to examine. The TOEFL Listening Test consists of three parts. In order to get a high score on the TOEFL Listening Test the examinees must understand 1) the meaning of sentences spoken, 2) the actual discourse between two speakers, and 3) lectures in a classroom setting.

However, the problem is the reliability of the external criterion listening test: the TOEFL Listening Test reliability is low ($r = 0.398$), a finding which calls for further research on the matter. Therefore, whether or not the Pronunciation Test might be eliminated remained unresolved because of the low reliability of the two tests. I will assume that a more reliable external criterion listening test could be found or tailored through in-depth item-test correlation analysis in future research.

GH 3, concerning the low reliability of the grammar tests, was supported in only Grammar Test A ($r = 0.099$). The reliability of Grammar Tests B ($r = 0.436$) and C ($r = 0.570$) was moderate. The low correlation between the Cloze Test and the Grammar Tests was supported as far as Grammar Tests A and B were concerned. There was a moderate correlation between the Cloze Test and Grammar Test C ($r = 0.441$, $p < 0.01$). In this study, however, it cannot be easily concluded that the Grammar Test C has far-reaching potential for becoming a reliable and valid grammar test. This is because in the three Grammar Tests different sentence structures and vocabulary are used. In the future, research on the effect

of test type or test format should be conducted in order to examine what kind of test types or test formats are appropriate to measure language learners' proficiency and the relationship with learners' proficiency levels in conditions in which the variable is limited to only test type or test format.

In addition, I must say that the lack of correlation in many places in the study may be due to the following three factors: 1) lack of reliability in the measures in general, and/or 2) lack of reliability with my specific group of students, and/or 3) restrictions in the range of ability that I sampled in the investigation. Tests are not simply reliable and valid in and of themselves. They are reliable and valid for specific types of students and specific ranges of ability (Brown 1983, 1996). Future research should be conducted, again under more controlled conditions, with a larger sample of participants with their proficiency levels taken into consideration.

Conclusion and Remaining Issues

In this study, I investigated the English questions used in university entrance examinations to determine if they are reliable and valid measures of examinees' English proficiency. The data indicated that, first, the NCUEE-Test is, to some degree, an appropriate measure of examinees' English proficiency in terms of reliability and validity. Second, there was no significant correlation between the TOEFL Listening Test and the Pronunciation Test. However, since both of the two tests showed low reliability, more research on this matter should be conducted in order to clarify whether the Pronunciation Test is or is not a valid measure of examinees' listening ability, measuring the correlation with a more reliable external criterion listening test. Third, test designers should try to produce more valid and reliable grammar tests. Further research is also needed to identify what kind of grammar tests have the potential of being reliable and valid testing devices. Moreover, this kind of research should be carried out again with a larger sample of participants, again with participants' proficiency level taken into consideration.

Acknowledgments

I sincerely thank Akihiko Mochizuki, Aichi University of Education; Shogo Miura and Tadashi Nishida, Hiroshima University; and J.D. Brown, University of Hawaii, for their valuable comments on earlier versions of this paper.

Akibiro Ito, Bachelor of Education, Aichi University of Education, is a M.A. student in the Department of English Language Education at Hiroshima University. His research interest is language testing, including the effects of test types on Japanese EFL learners' interlanguage performance.

Note

1. Klein-Braley and Raatz (1984) propose the criteria to judge the quality of the C-Test. In their six C-Test construction axioms, they say that a valid C-Test should correlate with a reliable discrete-point test at 0.5 or higher. Since C-Tests are one of the modified versions of cloze tests, I applied Klein-Braley and Raatz's idea for judging the concurrent validity of the Cloze Test in this study.

References

- Alderson, C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13, 219-227.
- Brown, J.D. (1983). A closer look at cloze: Validity and reliability. In J.W. Oller, Jr. (Ed.) *Issues in language testing research* (pp. 237-250). Rowley, MA: Newbury House.
- Brown, J.D. (1996). Personal communication.
- Brown, J.D., & Yamashita, S.O. (1995). English language entrance examination at Japanese universities: What do we know about them? *JALT Journal*, 17, 7-30.
- Daigaku nyushi series (University entrance exam series). (1994). Tokyo: Kyogakusha.
- Grammatik IV, version 1.0.2*. (1988). San Francisco: Soft Ware International.
- Ibe, T. (1983). Kyotsu ichiji shiken: Bunseki to kadai (PSCEE-Test: Analysis and problems). *Gendai Eigo Kyoiku (Modern English Teaching)*, 9, 6-7.
- Ikeura, S. (1990). Risuningu mondai wo kosatsu suru (Examination of listening tests). *Eigo Kyoiku (The English Teacher's Magazine)*, 11, 14-16.
- Ishii, K. (1981). Kyotsu ichiji no motarashitamono (What the PSCEE-Test has brought about). *Eigo Kyoiku Journal (Journal of English Teaching)*, 3, 9-13.
- Ishiguro, A. & G. Tucker (1989). Wang's story (I) & (II). *Short Stories for Reading Practice: Do you know...? 700-Word level*. Kyoto: Biseisha. 30-33.
- Kashima, H., A. Tanaka, Y. Tanabe, & K. Nakamura. (1983). *Zadankai: Daigaku nyushi eigo no mondatten* (Discussion: The problems with English tests in university entrance exams). *Gendai Eigo Kyoiku (Modern English Teaching)*, 8, 6-12.
- Kira, T. (1981). Kyotsu ichiji no meritto tsuikyuu wo (The pursuit of the PSCEE-Test's merits). *Eigo Kyoiku Journal (Journal of English Teaching)*, 3, 16-17.
- Kiyomura, T. (1989). *Dai roku sho: Eigo kyoiku hyokaron* (Chapter 6: Evaluation). In Y. Katayama, E. Endo., N. Kakita., A. Sasaki. (Eds.) *Shin eigo kyoiku no kenkyuu (New English language education research)* (pp. 244-246).

- Tokyo: Taishukan Shoten.
- Klein-Braley, C. & Raatz, U. (1984). A survey of research on the C-Test. *Language Testing*, 1, 134-146.
- Kuniyoshi, T. (1981). *Daigaku nyugaku shiken no kaikaku wa onset shiken no donyu kara* (The improvement of university entrance exams should begin with the administration of aural tests). *Eigo Kyoiku Journal (Journal of English Teaching)*, 3, 18-22.
- Masukawa, K. (1981). *Nippon eigo kyoiku katzen kondankai: Dai kyū kai taikai bokoku—Daigaku nyūshi no mondai wo chūshin ni shite* (The 9th round-table conference on the improvement of English education in Japan: A report with special reference to English questions in university entrance tests). *Eigo Kyoiku Journal (Journal of English Teaching)*, 3, 6-8.
- Ministry of Education, Science, and Culture. (1989). *Chugakko/kotogakko gakushu shido yoryo (Course of study for junior and high school)*. Tokyo: Japanese Ministry of Education, Science, and Culture.
- Mochizuki, A. (1984). Effectiveness of multiple-choice (M-C) cloze tests (2). *Research Bulletin*, 13, 159-164. Shizuoka: Chubu English Language Education Society.
- Mochizuki, A. (1992). Effectiveness of a multiple-choice (M-C) cloze test and the number of words in its text. *ARELE*, 3, 33-42. Hiroshima: The Federation of English Language Education Societies in Japan.
- Mochizuki, A. (1994). C-tests, four kinds of texts, their reliability and validity. *JALT Journal*, 16, 41-54.
- Negishi, M. (1990). Saikin no daigaku nyūshi mondai: Keiko to hyōka (Recent university entrance exams: Trends and evaluation). *Eigo Kyoiku (The English Teacher's Magazine)*, 11, 8-10.
- Nishida, T. (1985). *Kurouzu testo ni mirareru tekisuto no sakujo bensu ni suite* (Text and deletion variables in cloze tests). *Research Bulletin*, 15, 47-54. Hiroshima: Chugoku Academic Society of English Language Education.
- Nishida, T. (1986). Monogatariibun kurouzu to setsumeibun kurouzu: Tekisuto no sentaku to kurouzu testo (Narrative cloze and expository cloze: Text style and cloze tests). *Gengo Bunka Kenkyū (Foreign Language Cultural Research)*, 12, 124-143. Hiroshima: Department of Interdisciplinary Studies, Hiroshima University.
- Nishida, T. (1987). *Kurouzu test no datosei to testo no sentaku* (Cloze tests: Their validity and the selection of texts). *Eigo Kyōtōgaku Kenkyū (English Language Education Research)*, 433-444. Tokyo: Taishukan Shoten.
- Ohtomo, K. (1983). Kyōtsu ichiji shiken no eigo mondai (English questions in PSCEE-Tests). *Eigo Kyoiku (The English Teacher's Magazine)*, 8, 6-8.
- Sciarone, A.G., & Schoorl, J.J. (1989). The cloze test: Or why small isn't always beautiful. *Language Learning*, 39, 415-438.
- Shimizu, Y. (1989). *Eiken hikki shiken to cloze test ni mirareru soukan ni kansuru kenkyū* (A study on the correlations between subtests in the STEP Test and cloze tests). *STEP Bulletin*, 1, 103-116. Tokyo: Society for Testing English Proficiency.

- Shiozawa, T. (1983). *Datgaku nyushti: Genjo to kadai* (University entrance exams: State and problems). *Eigo Kyoiku (The English Teacher's Magazine)*, 3, 30-41.
- Steinberg, R. (1987). *Prentice Hall's practice tests for TOEFL*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Suzuki, H. (1981). Onsei testo donyu no kanosei (The possibility of administering aural tests). *Eigo Kyoiku Journal (Journal of English Teaching)*, 3, 23-25.
- Suzuki, S. (1985). Kyotsu ichiji shiken no mondaiten (Problems with the PSCEE-Test). *Gendai Eigo Kyoiku (Modern English Teaching)*, 11, 10-12.
- Takahashi, T. (1990). *Datgaku nyushti senta—shiken mondat* (University entrance exam center—NCUEE-Tests). *Eigo Kyoiku (English Education)*, 11, 11-13.
- Takahashi, Y. (1984). *Datzai no readability, gakushusha no kyomi, natyo no yasashisa to kurouzu tokuten no kanket ni tsuite* (The relationship between readability of tests, students' interests, easiness of contents, and the scores in cloze tests). *Research Bulletin*, 13, 14-22. Fukuoka: Kyushu English Language Education Society.
- Takahashi, Y. (1988). *Kurouzu test no bunmyaku izonset to tokaset* (The sensitivity and equivalency of cloze tests). *Research Bulletin*, 17, 33-40. Fukuoka: Kyushu English Language Education Society.
- Utsunomiya, M. (1985). *Koko nyushti kara datgaku nyushti e* (From high school entrance exams to university entrance exams). *Gendai Eigo Kyoiku (Modern English Teaching)*, 11, 18-19.
- Wakabayashi, S. & Negishi, M. (1990a). *Eigo testo mondatsakuset no tebiki* (English test questions (A guide for constructing English language tests). *Eigo Kyoiku (The English Teachers' Magazine)*, 5, 76-77.
- Wakabayashi, S. & Negishi, M. (1990b). *Eigo testo mondatsakusei no tebiki* (English test questions (A guide for constructing English language tests). *Eigo Kyoiku (The English Teachers' Magazine)*, 6, 78-79.
- Wakabayashi, S. & Negishi, M. (1991). *Eigo testo mondatsakuset no tebiki* (English test questions (A guide for constructing English language tests). *Eigo Kyoiku (The English Teachers' Magazine)*, 2, 38-39.
- Wakabayashi, S. & Negishi, M. (1994). *Musekinin na test ga ochikobore wo tsukuru* (Irresponsibly made tests keep students behind). Tokyo: Taishukan Shoten.
- Yanai, H., Maekawa, S., & Ikeda, O. (1989). Kyotsu ichiji gakuryoku shiken (Kokugo, Sugaku I, Eigo B) ni kansuru komokubunseki [Item analyses of PSCEE-Tests (Japanese, mathematics I, English B)]. *Datgaku nyushti forum (University Entrance Examination Forum)*, 11, 169-182.

(Received May 31, 1995; revised February 3, 1996)

Appendix: Cloze Test

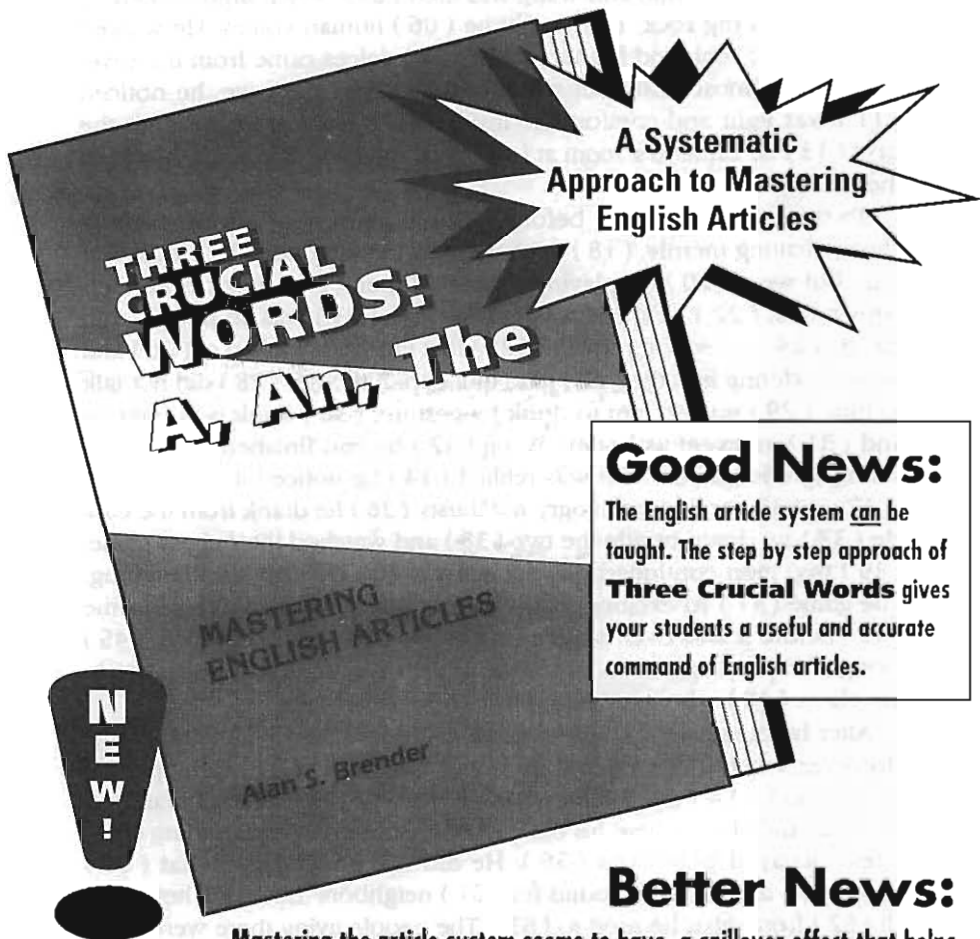
One day Wang lost his way while he was gathering wool. He wandered in the woods (01) hours, but could not find (02) path to lead him home. (03) came and Wang was tired (04) very hungry. When he passed (05) big rock, he thought he (06) human voices. He walked around (07) rock and found a cave. (08) voices came from the cave. (09) was almost dusk, but when (10) entered the cave, he noticed (11) was light and comfortable inside. (12) walked deeper into the cave (13) he came to a room at (14) end. Light and fresh air (15) from the ceiling.

Two men (16) sitting before a chess board. (17) were playing chess, chatting merrily. (18) neither talked to Wang nor (19) looked at him, but went (20) on playing. Now and then (21) drank from their cups which (22) held in their hands. Since (23) was so hungry and thirsty, (24) asked for a sip. For (25) first time they looked at (26) and smiled, offering him the (27) kind of a cup. Although (28) did not talk to him, (29) invited him to drink by gesture. (30) drink was fragrant and (31) as sweet as honey. Wang (32) he had finished it all, (33) strangely enough, the cup was refilled (34) he noticed it.

Wang (35) no longer hungry nor thirsty (36) he drank from the cup. He (37) sat down beside the two (38) and watched their chess game. (39) two men continued playing chess, (40) chatting and laughing. The game (41) so exciting that Wang became (42) in it. It took some (43) before it was over. Maybe (44) hour or more had passed, (45) thought. He had spent too (46) time in the cave, and (47) good-bye to the chess (48) who gave him a bag (49) a souvenir.

After he came out of (50) cave, he could find his (51) home easily. However, when (52) entered his home village and (53) some people on the road, (54) did not know any of (55). They were all strangers. He (56) the place where his old (57) was, but there was nothing (58) a few decayed poles and (59). He did not understand what (60) happened, and looked around for (61) neighbors' houses. They were all (62) from what he used to (63). The people living there were (64) strangers too. being at a (65) for what to do, he (66) the bag that the chess (67) had given him. Out came (68) stream of smoke, and in (69) minute, his hair had turned (70) and he found himself an old man. What does this story remind you of ?

THREE CRUCIAL WORDS, *The Definitive Guide to Mastering Articles in English*



**A Systematic
Approach to Mastering
English Articles**

**THREE
CRUCIAL
WORDS:
A, An, The**

**MASTERING
ENGLISH ARTICLES**

Alan S. Brender

NEW!

Good News:

The English article system can be taught. The step by step approach of **Three Crucial Words** gives your students a useful and accurate command of English articles.

Better News:

Mastering the article system seems to have a spillover effect that helps students with subject/verb agreement, noun number, use of possessives, use of demonstrative pronouns, correct personal pronoun references, proper word order, etc.

**HARCOURT
BRACE**

Harcourt Brace Japan, Inc.

Ichibancho Central Bldg, 22-1 Ichibancho, Chiyoda-ku, Tokyo 102

Tel: 81-3-3234-3912 ELT Hotline: 81-3-3328-2527 (Shari Berman)

Fax: 81-3-3265-7186

Fax: 81-3-3328-2213