مجله علوم اجتماعی و انسانی دانشگاه شیراز
دوره هیجدهم، شماره اول، تابستان ۱۳۸۱ (پیاپی ۳۵)
(ویژه نامه زبان انگلیسی و زبانشناسی)

دیکته تعبیری: روش جدید برای اندازه گیری درک شفاهی

دکتر حمید زاهدی
دانشگاه شیراز

چکیده

این مقاله به معرفی آزمونی جدید در حیطه درک شفاهی می پردازد. این روش که درک شفاهی کلان را اندازه‌گیری می کند دیکته تعبیری (دیکتو – فریز) نامیده شده است. در این روش، آزمودنی باید بر روی محتوای متنی که برای او خوانده می شود تمرکز کند و سپس واحدهای معنایی خواسته شده را به زبان و کلام خود بازسازی کند. بنابراین باز نویسی عین ساختار و لغات مربوطه متن اصلی مورد نظر نیست بلکه هدف بازسازی جان کلام می باشد. برای رسیدن به اهداف این مطالعه که تعیین روایی – پایایی و کارآیی دیکته تعبیری بود، سه آزمون درک شفاهی شامل تافل، نیم دیکته و دیکته تعبیری به ۷۴ آزمودنی داده شد. تجزیه و تحلیل نتایج نشان داد که دیکته تعبیری از روایی و پایایی قابل ملاحظه ای برخوردار است. علاوه بر این، سهولت ساخت، برگزاری و تصحیح دیکته تعبیری نشانگر کارآیی مناسب این روش می باشد.

JOURNAL OF SOCIAL SCIENCES & HUMANITIES
OF SHIRAZ UNIVERSITY
Vol. 18, No. 1, SUMMER 2002 (Ser. 35)
(Special Issue in English Language and Linguistics)

Test-Takers' Attitudes Toward The C-Test: The Effect of Practice

Dr. Hameed Zahedi*
Shiraz University

ABSTRACT

The main purpose of this study was to investigate whether the attitudes of test takers improve toward the C-test as a result of systematic practice over a period of 5 weeks. The sixty-three participants who were involved in the study had not heard of the procedure and, therefore, were taken to be impartial toward it. They were given a pretest followed by a questionnaire prior to the practice. Then followed a 5-week treatment after which a posttest was given and the same questionnaire was repeated to see whether any changes of attitude would take place. Although the participants' performance on the posttest improved significantly, their attitudes remained unchanged. The article, also, includes the answers to a number of other questions related to the C-test.

Key words: 1. C-test 2. Attitude

1. Introduction

1.1. Background

The C-test, which is a modification of the cloze procedure, was developed to compensate for the shortcomings of the cloze. One weakness of the systematic cloze, as stated by Raatz and Klein-Braley (1981), is that it does not guarantee the random sampling theory that is crucial for tests of reduced redundancy. Furthermore, Klein-Braley and Raatz (1984) express doubt about the validity of cloze tests. Moreover, Klein-Braley (1997) states that empirical findings show that the cloze is technically problematic with regard to issues like (a) the extreme length required in the classical cloze, (b) representativeness of individual cloze passages, (c) item bias, (d) scoring procedures, (e) failure of adult native speakers to make high scores, and (f) inter-dependency of items for the calculation of reliability. Besides, other researchers "see cloze as only measuring the ability to make localized connections in the text" (Alderson, 1978; 1979; 1983; Porter, 1983; Bachman, 1982; 1985, cited in Storey, 1997: 214).

Accordingly, as stated by Klein-Braley (1997: 63), "the C-test was an attempt to retain the positive aspects of cloze tests but to remedy their technical defects." Some of the improvements are: (a) the C-test is shorter than the cloze, (b) the deletion rate and start are fixed, and (c) the fixed scoring procedure ensures objectivity (Klein-Braley, 1997).

In the conclusion section of her paper, Klein-Braley (1997) discusses the superiority of the C-test over other test procedures in terms of difficulty level, reliability, validity, and factorial validity. However, she says little as to the attitude of the test takers toward the procedure, providing no empirical evidence for her statement that "students appear to accept the procedure as face valid, as they demonstrate by tackling the tests, even if they do not particularly enjoy doing them" (Klein-Braley, 1997: 71). However, the

* Assistant Professor of TEFL

present researcher does not think that 'tackling a test' could be an adequate criterion for determining the attitudes of test takers. As a matter of fact, face validity, as an aspect of attitude, is a matter of opinion or impression rather than tackling a test or performance. Face validity is "often determined impressionistically; for example, by asking students whether the exam was appropriate to their expectations" (Henning, 1987: 94). Thus, face validity, or test appeal, is the superficial judgment of the layman about the test appearance. However, attitude is more profound a concept, a multi-faceted notion which may not be stated adequately in vaguely general terms such as good or bad; it requires attention to a number of different dimensions which may be inter-correlated. Different aspects of attitude, as far as this paper is concerned, will be discussed below.

One investigation to find out about the attitude of the test takers toward the procedure was an attempt by Jafarpur (1995) who found that "C-tests do not possess face validity" (P. 209). To elaborate, he reported that 64% of the subjects' responses reflected disapproval of the procedure. They responded negatively to the questions in terms of whether the C-test was a test of English, a good test, a proficiency test, a fair test, a representative test, a complete test, a good entrance test, or a good test for the advanced. As a matter of fact, the participants rated the C-test as a spelling test, an IQ test, or a test of background-knowledge. Moreover, EFL instructors gave negative responses to 43% of the questions that may be interpreted as an indication of lack of confidence in the procedure among teachers. The author concluded his article asking whether this could be "the result of a lack of experience with the procedure on the part of the testes" (Jafarpur, 1995: 210).

1.2. Dimensions of Attitude

"Attitude can be considered a moderately intense emotion that prepares or predisposes an individual to respond consistently in a favorable or unfavorable manner when confronted with a particular object" (Anderson, 1990: 370). To operationalize the measurement of attitude, Kifer (1977) suggests a series of steps as follows: (1) Identification of the specific affective characteristic and appropriate target. (2) Delineation on the kinds of opinions that would be expressed by individuals possesses a positive attitude toward the target. (3) Identification of the dimensions of attitude toward the target. (4) Construction of sentences to measure the dimensions of the intended attitude.

Accordingly, to identify the attitude dimensions, important categories that are commonly used to specify the characteristics of test methods were selected. These are (a) face validity, (b) scalability validity, (c) discrimination power, (d) precision, (e) fairness, (f) seriousness, (g) effectiveness, (h) proficiency level, (i) difficulty level, (j) domain/skill appropriacy, (k) purpose, (l) popularity, and (m) trustworthiness. Furthermore, the categories were theoretically defined and grouped into six classes of attitude, as follows:

Attitude (1): It includes 7 categories, namely, face validity, scalability validity, discriminability, precision, fairness, seriousness, and effectiveness. The justification for classifying these categories into one group was that the same response mode was designed for their corresponding items. In other words, in order to answer to the questions that were developed on the basis of these categories, the participants were to choose among Yes, To some extent, No, and No idea. The response options for the other questions like 9 and 10 were different. Items 4 and 11 were each too detailed on their own to be included within Attitude (1). Here is a discussion of the theoretical definitions of the 7 categories:

(a) Face validity: "A test is said to have face validity if it looks as if it measures what it is supposed to measure" (Hughes, 1989: 27). Furthermore, Anastasi (1982: 136) defines

face validity as "the way the test appears superficially to measure" and adds, "it is not sufficient for a test to be objectively valid. It also needs face validity to function effectively in practical situations" (p. 136). Considerations for the presence of face validity is necessary since a face-invalid test may adversely affect test takers' performance (Hughes, 1989; Jafarpur, 1996). Bachman and Palmer (1996) do not consider face validity to be a separate quality of tests, arguing that it is essentially a function of authenticity and interactiveness.

(b) Scalability validity: This term refers to the capability of the test to identify "group differences" (Henning, 1987: 98) or differentiate among testees with different levels of proficiency. Thus, provided the test taker needs merely a very superficial knowledge of L2 to tackle test tasks is tantamount to saying that the procedure cannot "identify persons who qualify as competent and persons who do not qualify as competent" (Henning, 1987: 98). (c) Discriminability: While 'scalability' refers to the capability of the test to differentiate among various proficiency groups, 'discriminability' addresses the capability of the procedure to discriminate among test takers with heterogeneous abilities who belong to the same proficiency level. (d) Precision: It refers to the capability of the test to produce results that are reliable indicator of the test takers' ability. (e) Fairness: Fairness is concerned with whether the test produces results that are acceptable to test takers. The acceptability of the results may partly be an artifact of the nature of the test tasks. The more divergent the test tasks are from the purposes for which the test is used, the less fair the results and the decisions based on such results are perceived which in turn "can have serious consequences for test takers" (Bachman and Palmer, 1996: 32). Test method fairness, as far as the expected response is concerned, is also affected where individual's performance on one item affects his performance on other items. (f) Seriousness: "The 'bottom line' in any language testing situation, in a very practical sense, is whether test takers will take the test seriously enough to try their best, and whether test users will accept the test and find it useful" (Bachman, 1990: 288-9). Seriousness, as far as this study is concerned, corresponds with the degree of preparation for the test as demanded by test tasks. The degree of seriousness perceived by students may have a negative correlation with the degree of triviality of the test tasks. (g) Effectiveness: An effective test is capable of measuring test takers' ability efficiently with regard to the purpose of the test, producing meaningful results.

Attitude (2): (a) Proficiency level: This is to seek the test takers' attitudes as to which proficiency level they think the test is more appropriate for. In general, the more advanced the students are, the more sophisticated the testing method should be to allow eliciting higher order knowledge. (b) Difficulty level: Difficulty refers to how demanding the test tasks are judged to be. Generally speaking, performance on highly compact, or highly diffuse, input will be very demanding of test takers' competencies (Bachman, 1990), making processing too difficult.

Attitude (3): Domain/Skill appropriacy: It is concerned with the appropriateness of the test used in certain language domains. Here is a list of abilities, included in the questionnaire, which are, one way or another, related to the C-test: (a) general language ability, (b) reading comprehension, (c) writing, (d) grammar, (e) vocabulary, and (f) spelling.

Attitude (4): Function or purpose: "The appropriateness of a test is, to a larger extent, determined by purpose: why is a test needed at a certain stage and what use will be made of the result?" (Jafarpur, 1996). This section is concerned with the test takers' perceptions as to the functions that may appropriately be assessed by the C-test: (a) proficiency, (b) achievement, (c) entrance, (d) placement, (e) intelligence, and (f) aptitude.

Attitude (5): Popularity: "In a language test, test takers' affective schemata may influence the way in which they process and attempt to complete the test tasks.... Test performance can thus be facilitated or inhibited by positive or negative affective responses, both to the topical content of test tasks and to a particular type of test task" (Bachman and Palmer, 1996: 65-6). Accordingly, the students were given a list of seven common test methods to rank according to which they favored. These were (a) fill-in-the-blank, (b) essay-type, (c) multiple-choice, (d) true-false, (e) matching, (f) cloze-test, and (g) C-test.

Attitude (6): Trustworthiness: A feeling of mistrust of the test method fairness or appropriacy for a certain purpose has a negative impact on test takers' performance. While, each of the preceding individual items of the questionnaire (i.e., 1-12) attempted to elicit a specific aspect of the students' attitudes toward the procedure, Item 13 gets more personal, asking the students whether they have enough faith in the C-test to feel willing to have their language ability assessed by the procedure. In other words, do they believe that the C-test produces trustworthy results, which are representative of their abilities?

Following the identification and definition of the aspects of attitude, a question was developed to represent each attitude facet. The result was a modified Likert -scale questionnaire that has been discussed under Method below.

1. 3. The Present Study

Although it is commonly agreed that test takers' attitudes toward testing procedures have an impact on their performance, it seems that this area is under-researched. This problem multiplies for the C-test since the little research carried out on the attitude of test takers toward the procedure has brought about disagreement (see Klein-Braely and Raatz, 1984; 1997; Jafarpur, 1995). In fact, what was claimed by its proponents about the face validity of the C-test in 1997 was practically no different from what had been claimed, without attempting to provide empirical evidence, back in 1984: "The test ... is accepted by examinees as a legitimate language testing procedure, i.e. it also has face validity" (Klein-Braely and Raatz, 1984: 144).

Accordingly, the intent of this experiment was to seek an empirical answer to the question of attitude toward C-tests. Specifically, the investigator intended to seek an answer to the question of whether the subjects' attitudes toward this procedure improved as a result of practice. In the meantime, the researcher attempted to find out answers to the following questions: (a) Did the subjects' performance improve significantly as a result of practice? (b) Did the performance of male and female subjects on the C-tests differ significantly? (c) Did the attitudes of male subjects vary from those of the females in the two administrations of the questionnaire? (d) Did High, Mid, and Low groups, as determined by their scores on the post-administration of the C-tests, have comparable attitudes toward the procedure?

2. Method

2.1. Participants

The participants were 63 students, 40 male and 23 female, studying different branches of Engineering at Shiraz University. At the time of the study, they had a General English course with the investigator himself. The students were aged between 20 and 28.

2.2. Materials

The C-tests: Seventeen short passages were selected from Faces of the USA (Laird, 1987). They incorporated "lively and informative reading passages, suitable for students at intermediate level" (Laird, 1987). Accordingly, seventeen C-tests were developed

after deleting "every second word beginning from word two in sentence two" (Klein-Braely, 1997: 64). If the word included an odd number of letters then the larger "half" was deleted. Moreover,if the word had only one letter, then this word was ignored in the counting. The mean features of the developed C-tests have been displayed in Table 1.

Table 1. Average features of the 17 C-tests

Number of C-tests	Mean of words	Mean of mutilations	Mean of readability indexes based on Flesch Reading Ease
17	79.3	26.6	71.7

The C-tests were arranged from easy to difficult on the basis of their readability indexes. The 3 C-tests given in the first administration were the easiest with an average readability of 79.23. Except for the pre- and post-tests, which included 100 mutilations each, the C-tests given in each practice session included 50 mutilations. These exercises did not include 100 mutilations due to practical restrictions in terms of the class time, and also, the fact that the purpose was merely to give the subjects practice in taking C-tests.

The questionnaire: A questionnaire was developed to elicit the attitudes of the participants toward the C-test in terms of 13 categories, namely, face validity, scalability validity, discrimination power, precision, fairness, seriousness, effectiveness, proficiency level, difficulty level, domain/skill appropriacy, purpose, popularity, and trustworthiness. A question was constructed to represent each attitude facet.
A modified format of Likert -scale was used for the first 11 items. Item 12 was rank-based and Item 13 required limited production. Furthermore, an open-ended question was included (Item 14) to allow the participants make comments on other possible issues. Table 2 below displays the purpose of individual items of the questionnaire. The C-tests and the questionnaire have been reduplicated in Appendixes A and B, respectively.

Table 2: The functions of the items of the questionnaire

Items	Purpose of items	Items	Purpose of items
1	Face validity	8	Seriousness
2	Scalability validity	9	Proficiency level
3	Effectiveness	10	Difficulty level
4	Domain/skill appropriacy	11	Function or purpose
5	Fairness	12	Popularity
6	Precision	13	Trustworthiness
7	Discriminability	14	Free comment

2.3. Data Analysis

The C-tests: (1) Descriptive data: Basic descriptive statistics for the scores of the subjects on the pre- and post-tests were reported. (2) Administration-based comparisons: The paired t-test was utilized to determine possible significant changes in the performance of the subjects on the pre- and post-tests. (3) Gender-based comparisons: The independent sample t-test was used to determine possible significance differences between the performance of male and female subjects on the posttest, taking into account the fact that the number of male students was almost two times greater than that of the female.

The questionnaire: (4) Percentages: Percentage figures were provided on the responses of the participants on the two elicitation attempts for questions concerning proficiency level and difficulty level. (5) Administration-based comparisons: The number of responses to the items of the questionnaire from each person could serve as a

score for "overall attitude" (Henerson, Morris, and Fitz-Gibbon, 1987). Accordingly, The responses elicited via the questionnaire given after the pre- and post-tests were treated as individual's overall attitude scores. Then the paired t-test was employed to determine possible significant differences in the attitudes of the test takers toward the C-test. (6) Gender-based comparisons: The attitudes of the male subjects were compared against those of the female via the application of the independent sample t-test.

The interaction of C-tests with the questionnaire: (7) Score-Attitude comparisons: The participants were divided into three groups of High, Middle, and Low according to their performance on the posttest. Then, the performance of each was compared against the same group's attitudes toward the C-test to see whether the strength of performance was a determining factor in the attitudes of different groups.

2.4. Procedure

Pretest: In the pretest which included 3 C-tests with a total of 100 blanks, the students were allowed liberal time in which the slowest student took about 30 minutes to submit his/her test sheet. This was to assure that time was not a factor affecting the subjects' performance negatively. Immediately after the pretest, a 14-item questionnaire was given to the participants in their mother tongue. This was to elicit their attitudes towards the C-test.

Treatment: After the pretest, for a total number of 5 sessions--one session a week--the subjects received practice in doing C-tests before they were given the posttest. In each session, the subjects were given a couple of C-tests that included 50 mutilations. Based on pretest, the investigator set the time for these practice sessions to 20 minutes.

The researcher corrected the subjects' tests after each administration so as to find out the errors shared by most students. These included a variety of grammatical, lexical, semantic, textual, and spelling errors which were noted down to be used as a rich source of treatment to be given to the students right before the next administration in the next class meeting. A sample of the students' errors has been demonstrated in Table 3 below. Errors have been underlined and the correct word has been bracketed

Table 3. A sample of the subjects' errors

Errors	Explanation
Grammatical errors:	
At <u>this</u> (that) time	(this-that distinction)
coming <u>off</u> (of)	(of-off distinction)
At <u>these/those</u> (that) time	(agreement)
Textual errors:	
Some of them do have a lot of <u>monky</u> (money) to spend	(lexical cohesion: money-spend)
but usually they have earned <u>in</u> (it) themselves	(anaphoric: money-it)
They have love <u>them</u> (their) parents	(anaphoric: they-their)
Lexical and semantic errors:	
With feathers in their <u>hand</u> (hats)	
but <u>while</u> (white) people had better weapons	
where you can <u>withstand</u> (withdraw) money	
Spelling errors:	
<u>architectes</u>	
<u>competisions</u>	
<u>Citys</u>	

In the treatment, the researcher introduced strategies and made use of extensive exemplification. A sample of these instructions is listed below.

- It is more illuminating to move in a top-down (vs., bottom-up) fashion. Therefore, read the whole text first to acquire as much contextual information as you can.

- Likewise, it is more informative to work at sentence level (vs., word-level). So try to determine the beginning and end of the sentence first. Sentences are essentially made of noun phrases and verb phrases: S → NP VP.
- At the level of word, it is easier to start with re-building of function words (vs., content words). Once done with this, you can start from the beginning of the passage reconstructing the whole text.
- Blanks represent half (+1) of the letters of the whole word.
- NPs may require determiners. There are five types of determiners: Articles, Demonstratives, Possessives, Quantifiers, Numerals.
- NPs may require adjectives indicating the degree of the presence of certain attributes
- Modal or auxiliary verbs precede the main verb.
- Observe agreement!

Posttest: The posttest included 4 C-tests with a total number of 100 mutilations and liberal timing as that of the pretest. The same questionnaire was given to the participants after the posttest to find out possible change in their attitudes toward The C-test.

3. Results

The C-tests: Table 4 gives the basic statistics for the scores of the participants on the pre- and post-tests.

Table 4: Basic descriptive statistics for the scores on pre- and post-tests

Variable	Mean	SD	Min (0)	Max (100)
Pre-test	41.92	22.48	1	90
Post-test	47.43	15.68	8	95

The pretest mean was 41.92, whereas the posttest's stood at 47.43. As the post-test average showed improvement, the dispersion of the scores narrowed down. Table 5 displays the results of the paired t-test that was utilized to compare the participants' performance on the pre- and post-tests.

Table 5: Paired t-test for the pre- and post-tests

Variable	Pairs	Corr.	2-tailed sig.	t-value	df	2-tailed sig.
Pre- and Post-tests	63	0.81	0.000	3.24	62	0.002

The t-value was 3.24 that was highly significant at a probability level of 0.002. Also, the correlation coefficient between the two variables was 0.81 that was of high significance. Table 6 shows the results of the independent-sample t-test employed to find out whether there were meaningful differences between the performance of male and female participants on the post-test.

Table 6: Independent-sample t-test for gender differences on the post-test

Variable	Cases	Mean	SD	t-value	df	2-tailed sig.
Female	23	45.48	14.16			
				– 0.75	61	0.46
Male	40	48.55	16.55			

The male students almost doubled the female. The male average was slightly larger than that of the female; however, the t-value, which stood at – 0.75, indicated that the difference may be attributed to chance.

The questionnaire:

Attitude (1): The attitude scores of individual students on each dimension of Attitude 1 were added to arrive at each individual's overall attitude score. Positive responses were assigned 3 points; relative certainties were given 2 points; negative answers were assigned 1 point; furthermore, the abstained cases were given no points.

Therefore, the overall attitude score of each individual could possibly range between 7 to 21 representing absolute disapproval and approval respectively. This range has been shown below in Figure 1.

Figure 1: Attitude scores continuum (individual-based)

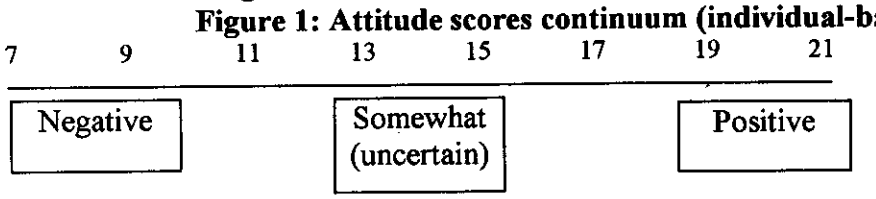

Table 7 demonstrates the results of the paired t-test that was applied to see whether there were any meaningful differences between the attitude scores of the students on the first and second administrations of the questionnaire.

Table 7: Paired t-test for pre- and post-attitudes on Attitude 1

Variable	Pairs	Correlation	2-tailed sig.	Mean	SD	t-value	df	2-tailed sig.
Pre-attitude				10.89	2.65			
	63	0.59	0.000			1.70	62	0.95
Post-attitude				11.43	2.90			

The means of the pre- and post-attitudes were 10.89 and 11.43 respectively. The difference between the averages was too small to reveal any significant meaning. However, there was a meaningful correlation between the two variables with an index of 0.59.

Attitude (2): Tables 8 and 9 display how the students thought of the proficiency level and difficulty level of the C-tests.

Table 8: Pre- and post-attitudes on the proficiency level of C-tests

Attitude	Beginning	Intermediate	Advanced
Pre-attitude	5 (8%)	18 (29%)	40 (63%)
Post-attitude	4 (6%)	26 (42%)	33 (52%)

Table 9: Pre- and post-attitudes on the difficulty level of C-tests

Attitude	Easy	Somewhat difficult	Difficult
Pre-attitude	0 (0%)	25 (40%)	38 (60%)
Post-attitude	4 (6%)	28 (45%)	31 (49%)

On the whole, it seems that a great majority of the students thought of the C-tests as (somewhat) difficult, more appropriate for intermediate or advanced levels of proficiency.

Attitude (3): Table 10 shows the students' pre- and post-attitudes as to the appropriacy of the C-test for different domains of language use. Based on Figure 1, YES, SOMEWHAT, and NO responses were assigned 3, 2, and 1 points respectively to calculate the overall mean attitude for each skill. The possible attitude scores for the two extremes were 63 x 3 = 189 for perfect approval and 63 x1 = 63 for perfect disapproval with the midpoint at 63 x 2 = 126. Thus the total attitude scores above 126 tended to go for the appropriacy of the C-test for the intended skill, and vice versa. Figure 2 displays the relationship.

Figure 2: Attitude scores continuum (item-based)

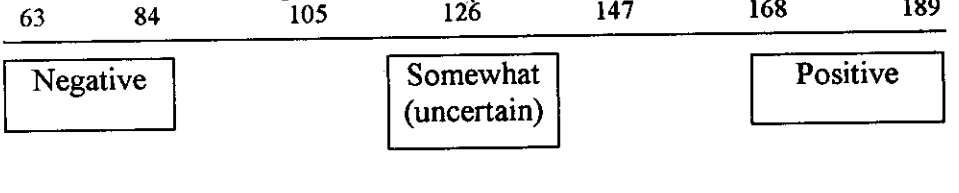

Table 10: Pre- and post-attitudes on domain/skill appropriacy of C-tests (in number) as well as total and mean attitude scores

Attitudes	Yes	Some-what	No	Total attit. scores	Mean attit. scores
Skills/Domains	Pre. (Post.)	Pre. (Post.)	Pre. (Post.)	Pre. (Post.)	Pre. (Post.)
Gen. Lang. Ability	24 (27)	30 (28)	9 (8)	141 (145)	47.00 (48.33)
Reading comp.	34 (27)	22 (30)	7 (6)	153 (147)	51.00 (49.00)
Writing	29 (30)	22 (24)	12 (9)	143 (147)	47.67 (49.00)
Grammar	14 (25)	30 (27)	19 (11)	121 (140)	40.33 (46.67)
Vocabulary	40 (39)	13 (17)	10 (7)	156 (158)	52.00 (52.67)
Spelling	26 (28)	19 (18)	18 (17)	134 (137)	44.67 (45.67)

According to the results, grammar and spelling had the lowest scores, whereas vocabulary won the highest score in both pre- and post-attitudes. Analysis of Variance followed to see whether the students preferred the C-test to be used in certain domains.

Table 11: ANOVA for preferences of the C-test to certain domain(s) based on pre-attitude scores

Source of variance	df	SS	MS	F-ratio	Significance
Between groups	5	273.78	54.75	0.0338	0.9992
Within groups	12	19432.00	1619.33		
Total	17	19705.78			

Table 12: ANOVA for preferences of the C-test to certain domain(s) based on post-attitude scores

Source of variance	df	SS	MS	F-ratio	Significance
Between groups	5	87.78	17.55	0.0105	1.0000
Within groups	12	20114.67	1676.22		
Total	17	20202.44			

The F-ratios standing at 0.03 and 0.01 for the pre- and post-attitudes, as depicted in Tables 11 and 12 above, showed no meaningfully specific preferences for any skills. In order to capture significant differences in the attitudes of the students, with regard to skill/domain appropriacy of the C-test, in the two administration of the questionnaire paired t-test was employed. The results have been presented in Table 13 below.

Table 13: Paired t-test for pre- and post-attitudes on Attitude 3

Variable	Pairs	Correlation	2-tailed sig.	Mean	SD	t-value	df	2-tailed sig.
Pre-attitude				47.11	34.05			
	18	0.95	0.000			–0.55	17	0.59
Post-attitude				48.55	34.47			

The t-value was –0.55 that was statistically insignificant. However, the association between the two variables was highly significant with an index of 0.95.

Attitude (4): Table 14 depicts the pre- and post-attitudes of the 63 participants about the purpose(s) for which the C-test may be appropriately used. The procedure of scoring and reporting the results of Attitude 4 is the same as that of Attitude 3. Accordingly, the total attitude scores above 126 reflected positive attitudes toward the C-test for the intended purpose(s), and vice versa (see also Figure 2 above).

Table 14: Pre- and post-attitudes on the purpose of the C-test (in number) as well as total and mean attitude scores

Attitudes	YES	SOMEWHAT	NO	Total attit. scores	Mean attit. scores
Purposes	Pre. (Post.)	Pre. (Post.)	Pre. (Post.)	Pre. (Post.)	Pre. (Post.)
Proficiency	15 (20)	27 (33)	21 (10)	120 (136)	40.00 (45.33)
Achievement	12 (10)	22 (25)	29 (28)	109 (108)	36.33 (36.00)
Entrance	7 (6)	15 (11)	41 (46)	92 (86)	30.67 (28.67)
Placement	15 (15)	25 (31)	23 (17)	118 (124)	39.33 (41.33)
Intelligence	25 (19)	25 (30)	13 (14)	138 (131)	46.00 (43.67)
Aptitude	32 (30)	27 (23)	4 (10)	154 (146)	51.33 (48.67)

According to the results, entrance had the lowest scores, whereas aptitude got the highest score in both pre- and post-attitudes. Then the Analysis of Variance was employed to see whether the students preferred any specific purpose(s) for the C-test. The results have been demonstrated in Tables 15 and 16 below.

Table 15: ANOVA for preferences of the C-test to certain purpose(s) based on pre-attitude scores

Source of variance	df	SS	MS	F-ratio	Significance
Between groups	5	789.61	157.92	0.2528	0.9303
Within groups	12	7496.67	624.72		
Total	17	8286.28			

Table 16: ANOVA for preferences of the C-test to certain purpose(s) based on post-attitude scores

Source of variance	Df	SS	MS	F-ratio	Significance
Between groups	5	782.94	156.59	0.2288	0.9427
Within groups	12	8213.33	684.44		
Total	17	8996.28			

The findings showed no meaningful preferences to any of the purposes with F-ratios standing at 0.25 and 0.23 for the pre- and post-attitudes, respectively. Then to find possible changes in attitudes in the two administrations of the questionnaire, a paired t-test was applied. The results have been demonstrated in Table 17 below.

Table 17: Paired t-test for pre- and post-attitudes on Attitude 4

Variable	Pairs	Corr.	2-tailed sig.	Mean	SD	t-value	df	2-tailed sig.
Pre-attitude				40.61	22.078			
	18	0.92	0.000			0.00	17	1.00
Post-attitude				40.61	23.004			

The t-value was 0.00 revealing no change at all. However, the go-togetherness of the two variables was highly significant with a coefficient of 0.92.

Attitude (5): Table 18 shows the testees' attitudes toward the tests as elicited in the first and second administrations. They have been arranged according to their popularity.

Table 18: Popularity rank based on pre- and post-attitudes

Test method	Pre-attitude mean scores	Post-attitude mean scores	Popularity rank
Multiple-choice	1.83	1.76	1
True-False	3.22	3.25	2
Matching	3.68	3.63	3
Fill-in-the-blank	3.71	3.79	4
Cloze-test	4.88	4.95	5
Essay-type	5.14	5.04	6
C-test	5.58	5.63	7

The C-test stood last in popularity both in the pre- and post-administrations of the questionnaire, whereas the multiple-choice format occupied the first place. The correlation index for the pre- and post-attitudes was almost *perfect (0.99).*

Attitude (6): Almost 50% of the students did not express inclination to have their language ability assessed via the C-test. The other 50% mentioned their priorities as follows: (a) Skills: vocabulary 16, general language ability 13, reading comprehension 12, grammar 8, writing 7 spelling 4, (b) Purpose: aptitude 14, proficiency 12, achievement 10, Intelligence 8, placement 4, entrance 0

Gender-based comparisons: Tables 19 and 20 give the results of the independent-sample t-tests employed to find out possible significant differences between the attitude scores of male and female students on Attitude 1 in both administrations of the questionnaire.

Table 19: Independent-sample t-test for gender-based comparisons on pre-attitude scores

Variable	Cases	Mean	SD	t-value	df	2-tailed sig.
Female	23	10.09	1.88			
				– 1.86	61	0.068
Male	40	11.35	2.93			

Table 20: Independent-sample t-test for gender-based comparisons on post-attitude scores

Variable	Cases	Mean	SD	t-value	df	2-tailed sig.
Female	23	10.52	2.63			
				– 1.92	61	0.059
Male	40	11.95	2.95			

Although male students were more favorable toward the C-test than the female, the t-values, –1.86 and – 1.92, indicated that the differences were by chance.

Score-Attitude comparisons: Table 21 summarizes the basic statistics belonging to high, mid, and low groups based on the post-administration of the C-test. The high group included students scoring at and above 55; the range of the mid-group was between 40 and 54; the low group consisted of students scoring below 40.

Table 21: Basic descriptive statistics for the scores of High, Middle, and Low groups

Groups	Mean	SD	Minimum (0)	Maximum (100)
High	64.68	12.15	55	95
Middle	46.56	3.82	40	53
Low	29.53	8.21	8	39

As a matter of fact, the borderlines between categories that were drawn judgmentally were proven to be valid. The F-ratio was highly significant standing at 81.14. Table 22 shows the results. The F-ratio was then committed to a Scheffe test to find the whereabouts of the differences. The test indicated significant differences among all the three groups.

Table 22: ANOVA for High, Middle, and Low groups' performance differences

Source of variance	df	SS	MS	F-ratio	Significance
Between groups	2	11124.42	5562.21	81.1408	.0000
Within groups	60	4113.01	68.55		
Total	62	15237.43			

Table 23 depicts the basic attitude scores of the high, middle, and low groups on Attitude 1 based on the post-administration of the questionnaire.

Table 23: Basic statistics for the attitude scores of High, Middle, and Low groups

Groups	Mean	SD	Minimum (7)	Maximum (21)
High	13.21	2.90	8	18
Middle	10.93	2.61	7	17
Low	10.23	2.51	7	16

The high group showed greater interest in the C-test as compared with the other two

groups although their attitude mean score, 13.21, was still lower than 14--the midpoint between the perfect positive and perfect negative attitude scores as shown in Figure 1 above. Moreover, ANOVA was used to capture meaningful differences between the attitude means.

Table 24: ANOVA for High, Middle, and Low groups' attitude differences

Source of variance	df	SS	MS	F-ratio	Significance
Between groups	2	81.00	40.50	5.17	0.0085
Within groups	60	469.86	7.83		
Total	62	550.86			

As shown in Table 24 above, the F-ratio was highly significant at 5.17. However, the Scheffe test revealed differences between the high group with the mid-group and the low group only. That is, there were no meaningful differences between the two latter groups.

4. Discussion and Conclusion

4.1. Discussion

Main question:

Although their performance on the C-tests improved significantly (Tables 5 and 6), the participants' attitudes remained unchanged (Tables 8-19). However, the association between the pre- and post-attitudes was always highly significant.

Other questions: (a) The test takers performance on the post-test improved meaningfully (Tables 4 and 5). The pre-test mean that was 41.92 turned into 47.43 after 5 practice sessions. That is to say that the treatment was statistically effective. Messick (1996: 246) states that "test preparation practices emphasizing test familiarization and anxiety reduction may actually improve validity: scores that formerly were in validity low because of anxiety might now become validly higher." However, as a matter of fact, the improvement was not so great because the post-test results brought about 5.5% gains in score. In other words, the distance between the mean, 47.43, and the perfect score, 100, was still great.

(b) The performance of male and female students on the post-test administration of the C-tests did not differ significantly (Table 6) although the male performed better with a mean score of 45.48 versus that of the female standing at 45.48.

(c) The attitudes of male students did not vary meaningfully from those of the female in both administrations of the questionnaire (Tables 19 and 20) although the male were less negative toward the procedure as opposed to the female.

(d) The high, middle, and low groups had comparable attitudes toward the procedure (Table 23). The high group's overall attitude was significantly different from both the middle and low groups'. However, middle and low groups' differences of attitudes were not meaningful (Table 24).

4.2. Conclusion

1) It was rather difficult in the first place to interpret the students' attitudes as to proficiency level, difficulty level, and seriousness of the C-test. Compare the figures in Table 25.

Table 25: Attitudes to proficiency level, difficulty level, and seriousness of C-test

Proficiency level	Beginning (6%)	Intermediate (42%)	Advanced (52%)
Difficulty level	Easy (6%)	Somewhat (45%)	Difficult (49%)
Seriousness	Yes (0.14)	Somewhat (0.20)	No (0.66)

While the majority of the students thought of the C-test as (somewhat) difficult, appropriate for intermediate and advanced levels, they did not take the procedure seriously. This seems to be paradoxical since difficult and advanced tasks are expected to make the students take the tests more seriously. The evidence for the difficulty of the

C-test comes from the performance of the students that was below 50% of the perfect, on the average. Moreover, many of the students stated that the test was too demanding in terms of language competence, hence the appropriacy of the test for intermediate and advanced levels of proficiency. The answer to this paradox may be that it was due to the very face of the test that most of the students did not think of it as a serious test of English. In fact, they tended to think of the C-test as a difficult test measuring trivial points that do not require extensive preparation before the test. Another interpretation may be that it is difficult to know what and how to prepare before the C-test. This latter interpretation is evidenced by the responses of the participants to the items concerning domain and purpose of the C-test that seemed to be vague for almost all of the students (Tables 10-12 and 14-16).

2) The differences between attitudes of the students toward skill/domain appropriacy of the C-test indicated that they were undecided about the idea. This was supported by the fact that the students scored different domains with doubt. The total scores ranged from 121 to 156 (Table 10) that fell around the center of the continuum in Figure 2. Moreover, the attitude scores for different domains were not significantly different (Tables 11 and 12). This is to say that the C-test enjoyed the same degree of appropriateness for all of the skills.

3) The same problem, as in 2 above, was observed for the purpose of The C-test. The total attitude scores ranged from 92 to 154 (Table 14) that fell around the mid-point of the continuum in Figure 2. Furthermore, the attitude scores for different purposes were not significantly different (Tables 15 and 16). This amounted to saying that the C-test may be, more or less, utilized (in fact, somewhat) for all of the purposes.

4) It seemed that selected-response test methods, involving test takers in recognition, selection, and matching, enjoyed the highest popularity. Next came fill-in-the-blank and the cloze that involve limited production, followed by the familiar essay-type. The unfamiliar C-test, which was unusually production-based, came last for several reasons stated by almost half of the test takers: (a) too demanding in terms of language competence and/or intelligence, (b) not effective, (c) too many mutilations and so too difficult, (d) no fair results due to item dependency and exact scoring procedure, (e) too narrow in terms of assessing capabilities, (f) not face valid, (g) vague purpose, (h) unfamiliar format, (i) inappropriate for other majors (vs., English majors), (j) inappropriate for testing purposes (vs., practicing purposes and class exercises), (k) producing irritation

In sum, the overall attitude of the students did not improve as a result of practice, and the test was shown to have low "public acceptability" (Weir, 1990: 49) both before and after the students' familiarization with the test.

References

Alderson, J. C. (1978). *A Study of the Cloze Procedure with Native and Non-Native Speakers of English,* Unpublished Ph.D Thesis, University of Edinburgh.

Alderson, J. C. (1979). *The Cloze Procedure and Proficiency in English as a Foreign Language,* **TESOL Quarterly,** 13, 219-27.

Alderson, J. C. (1983). *The Cloze Procedure and Proficiency in English as a Foreign Language,* In Oller, J. W. Jr. (Ed.), **Issues in Language Testing**, Rowley, MA: Newbury House, 219-28.

Anastasi, A. (1982). **Psychological Testing,** London: McMillan.

Anderson, L. W. (1990). *Attitudes and their Measurement,* In Walberg, H. J. and

Haertel, G. D. (Editors), **The International Encyclopedia of Educational Evaluation,** Oxford: Pergamon Press, 368-374.

Bachman, L. F. (1982). *The Trait Structure of Cloze Test Scores,* **TESOL Quarterly** 16, 61-70.

Bachman, L. F. (1990). **Fundamental Consideration in Language Testing,** Oxford: Oxford University Press.

Bachman, L. F. (1985). *Performance on Cloze Tests with Fixed Ratio and Rational Deletions,* **TESOL Quarterly,** 19, 535-56.

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

Henerson, M. E., Morris, L. L., and Fitz-Gibbon, C. T. (1987). **How to Measure Attitudes,** Newbury Park, California: Sage Publications.

Henning, G. (1987). **A Guide to Language Testing**: **Development, Evaluation, Research,** Cambridge: Newbury House Publishers.

Hughes, A. (1989). **Testing for Language Teachers,** Cambridge: Cambridge University Press.

Jafarpur, A. (1995). *Is C-testing Superior to Cloze?* **Language Testing,** 12(2), 194-216.

Jafarpur, A. (1996). **A Course in Language Testing,** Tehran: Payame Noor University.

Kifer, E. (1977). *An Approach to the Construction of Affective Evaluation Instruments,* **Journal of Youth Adolescence,** 6, 205-14.

Klein-Braley, C. (1997). *C-Tests in the Context of Reduced Redundancy Testing: An Appraisal,* **Language Testing,** 14(1), 47-84.

Klein-Braley, C. and Raatz, U. (1984). *A Survey of Research on the C-test,* **Language Testing,** 1, 134-46.

Laird, E. (1987). **Faces of the USA,** New York: Longman.

Messick, S. (1996). *Validity and Washback in Language Testing,* **Language Testing,** 13 (3), 241-256.

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

Porter, D. (1983). *The Effects of Quality of Context on the Ability to Make Linguistic Predictions: A Flaw in a Measure of General Proficiency,* In Hughes, A. and Porter, D. (Eds.), **Current in Language Testing,** London: Academic Press, 63-74.

Raatz, U. and Klein-Braley, C. (1981). *The C-Test--A Modification of the Cloze Procedure,* In Culhane, T., Klein-Braley, C., and Stevenson, D. K. (Eds.), **Practice and Problems in Language Testing,** University of Essex Occasional Papers 26. Colchester: University of Essex.

Richards, J. C., and Platt, J., and Platt, H. (1992). **Longman Dictionary of Language Teaching and Applied Linguistics**, London: Longman.

Storey, P. (1997). *Examining the Test-Taking Process: A Cognitive Perspective on the Discourse Cloze Test,* **Language Testing** , 14 (2), 214-231.

Spolsky, B. (1985). *The Limits of Authenticity in Language Testing,* **Language Testing,** 2 (1), 31-40.

Weir, C. J. (1990). **Communicative Language Testing**, London: Prentice Hall.

APPENDIX A: The C-tests

1. BUILDING INDUSTRY, P. 50

America's oldest cities, such as Boston, San Francisco, and New York, were built before the days of the automobile. At th------ time, ever----- had t------ live cl------ to t---- city cen-------. But t------ coming o---the c------- changed a------- that. Mo------- and mo-------of th------- who co------- afford t------- buy th------- own ho------- moved o------ - to t------- new sub-------, where th------- could for------- the no-------, dirt, a------- crime o------- the ci-------

2. TRANSPORTATION, P. 58

Americans love speed. They li------ to co------ in micr------ ovens, Th------- prefer mak------- phone ca----- to wri------- letters, a------- they li------- to tra------- by a------- rather th------- bus o------- train. A------- when th------- send a pac------, they wa------ it t------ arrive tomo------, not ne------ week. I------- a cou------ the si---- of t------- U.S., th------ is n------- easy. Dist------- are enor-------, and wi------- the sp------- limit o------- most high------- at on------- 55 mph , i------- takes a lo------- time f------- things t------- go b------- land. Everything that can fly, does.

3. CHANGING INDUSTRIES, P. 60

A young man with no money and little education decides he must get to the top. He wo------ all d-----, studies a------- night, a------- looks aro------- for a cha------ to g------ rich. H------- discovers th------- all h---- friends li------ eating h------- mother's home-ma------- pizzas. H------- learns h------ to ma------ them a------- starts sel-------them a------ school. H------- leaves h----- job a------ opens a pi----- restaurant. I---- is a gr--- success, s----- he op----- another, th------- another. Fi------- years la-------, he h------- made h----- first mil---- dollars. H------- is t------- perfect exa------- of the successful American businessman.

4. BUILDING INDUSTRY, P. 50

In 1871, Mrs. O'Leary's cow kicked over a lantern and set fire to Chicago. Eighteen thou---- buildings bur------- down, a------- three hun------- people di-------. It w------- the e------ of t------ old Chi------, but i------- was al------ a n------ beginning. O------ of t------ ashes gr------ a wh----- new st------ of buil------. The archi---- who reb------- Chicago us------- steel t------ make str----- frames th------- climbed s------- high th------- scraped the sky. The skyscraper was invented.

5. YOUNG PEOPLE, P. 8

It is not always easy to keep a job, and still do well in school. But Amer----- children le------ early th------ you ha------- to wo------- hard t------- win. Win-------, in fa------, is pa----- of t------ American w----- of li------. Americans ju------ like compet------, even i---- their ti------ off. F------ people c------- enjoy a sp------- unless th------- are try------- to be------- someone el------- at i-------.

6. THE FIRST AMERICANS, P. 6

You probably know the romantic names of the most famous American Indian tribes: Apache, Sioux, Cherokee, and Comanche. They prob------- make y------- think o------- men rid------- horses wi------- feathers i------- their ha-------, and wo------- with lo-------, plaited ha------- and so------ leather dre------. But th------ are, i------- fact, hund------- of tri------- of Ind-------, all wi------- different cus------. Some ha------ always li------ in hou------ and gr------ their o------ food. So------- used to move from place to place and hunt animals. Some were fishermen, living by the sea.

7. HEALTH AND FITNESS, P. 78

"Americans are not as fit as they think they are," says Dr. Michael McGinnis of the U.S. Department of Health and Human Services. That sho------ come a------ a surp-----, because i------- the Sta------, fitness i------ something y------ cannot g------ away fr-------

You wi------ see spe------ stores sel------- clothes f------- exercise. Y------- will fi------- the sto------- full o------- books a------- cassettes tel------- you h------- to g------- fit. Y------ will me------ joggers i------ the pa------, and find at least one health club in every town.

8. YOUNG PEOPLE, P. 8

Typical American teenagers are in fact very ordinary. They th----- their teac----- make th----- work t------ hard, th------ love th------- parents b------- are su------- they d------ not under------- anything, a------- their frien------ are t------ most impo------ things i------ their li------- Some o------- them d------- have a l------- of mo------- to sp------, but usu------ they ha------ earned i------ themselves. Mo------ young peo------ take jo------- while th------- are i------- school. They work at movie theaters, fast food restaurants, gas stations, and stores to pay for their clothes and entertainment. Maybe this is what makes them so independent from their parents at such a young age.

9. RICH AND POOR, P. 14

Most Americans say they are better off now than they were five years ago. However, i----- recent ye-----, more a------- more peo------ have bec------- trapped i------- an under-------. Many, b------- not a----- of th-----, are bl------. Many, b------- not a-------, live i------- the o------- inner cit-------. These peo------- seem t------- be una------- to esc------ from b------- housing, unemployment, and a life of crime and hopelessness. For them drugs and alcohol are specially serious problems.

10. THE ROCKY MOUNTAINS, P. 24

The early European settlers in America lived on the East Coast. To the west, beyond a line they called the frontier, the land was full of unknown dangers. The pion------- pushed fur------- and fur------ west un------- they rea------- the Pac------- Ocean a ------- the fron------- was th------- no lon------. But t------- idea o------- the la------- beyond t------ frontier, t------ land wh------ a m------- could g------- and ta------- what h------- wanted wi------- his sk------- and cou------- had bec------- part of the American dream. The results of this dream can be seen most clearly in the Rockies, the great mass of mountains that runs down the western side of the States.

11. THE FIRST AMERICANS, P. 6

When the Europeans came, the Indians watched with horror as the forests were cut down, the prairie grasses ploughed up, and the wild animals destroyed. They thems------ were slo------ driven o----- the go----- land. Some------ they fou------, but wh------ people h----- better wea----- and alw----- won i----- the e------. Sometimes th------ made agree------, but t------- white m------- always br------- his prom-------. Some tri------- died o------- completely a------ were lo------- forever. T------- American Ind------- had fin------- been bea-------.

12. THE AUTOMOBILE INDUSTRY, P. 56

Today, in America, only really poor families and those too old to drive do not own a motor vehicle. And th------- freedom i-------- limited. B------- for t------ 87 per------- who d------- have ca------, there i------ hardly a------ need t------ use th------- legs. Th------ are ba-------, fast-fo------- restaurants, a------- movie thea-------- where y------- can with------- money, e------- a me-------, or s------- a fi------- without ev------- getting o------ of y------- car. Th------- are ev------- drive-in churches. It's surprising that some people remember how to walk at all.

13. EDUCATION, P. 40

Americans have always believed in education , but in a specially American way. The scho---- first j------- was t------- turn mill------- of for------ chilin------ Americans. Si------ they ca------- from doz------- of diff------- countries, th----- was n------ easy. Sch------ had t------ teach t------ children t------- speak Eng-------, to lo------- their n------- country, a------- to le------- how t------- live i------- it.

14. BUILDING INDUSTRY, P. 50

The idea of building skyscrapers soon spread to other cities. Manhattan i------- New Yo------, became a for------- of skysc-------. All ov------- the cou-------, city skyl------- changed, a------- banks, b------- companies, a------- hotels pus------- concrete, gl-------, and st------ higher a------ higher. I------- downtown ar-------, where t------- stores a------- offices we------- found, ear------- buildings disap------- to ma------- way f------- the n------- style.

15. THE AUTOMOBILE INDUSTRY, P. 56

The car is such an important part of American life that for many people it would be impossible to manage without it. A home------- living i------- a sub------, for exa-------, probably h------ a twenty-min----- drive t------ take h------- children t------- school. S------ then tu------- the c------- around a------- drives f------- half a------- hour i------- another dire------ to g------ to h------- job i------ an off------. To d------ her shop-------, she h------- another lo------- drive, s------- she plans carefully and buys food for two weeks in one trip.

16. BUILDING INDUSTRY, P. 50

As a matter of fact, different areas of the American cities do not develop equally. In ot----- words, wh----- the down----- areas o------ many cit------ become mo---- modern, ot----- parts a----- falling in------ ruin. I------ the in-------city ar----- poorer peo------- are cro------ into a------badly cared-f------ buildings. Th------ are t------ few jo-------, and th------ is t------ much cr------, and t------ little ho------.

17. TRANSPORTATION, P. 58

Offering a quick service is a sure way to make money in the States. Fast-delivery comp----- are get----- more busi------ all t------- time. O------- of t------- biggest o------- them, Fed-------- Express, wi------- deliver a comp------ disk i------ just a f-------hours t------ a forg------- businessman w------- left h------- bag beh------. Or i------ will g------ a tux------ to a col----- student away fr------ home w------ has a------ unexpected da------- and not------- to wear.

APPENDIX B: The Questionnaire

آزمونی که هم اکنون در آن شرکت کردید سی- تست نام دارد. لطفا به سؤالات پرسشنامه زیر که مربوطه به این آزمون مـی باشـد پاسخ دهید. زیر کلمات کلیدی جهت تأکید بیشتر خط کشیده شده است.

۱- در نظر اول، آیا شما فکر می کنید ظاهر این آزمون به یک تست زبان انگلیسی شباهت دارد؟

بلی ☺ تا اندازه ای 😐 خیر ☹ نظری ندارم ○

۲- آیا کمی آشنایی با زبان انگلیسی می تواند برای پاسخگویی به این تست کافی باشد؟

بلی ☺ تا اندازه ای 😐 خیر ☹ نظری ندارم ○

۳- آیا شما این آزمون را بعنوان یک تست مؤثر زبان انگلیسی ارزیابی می کنید؟

بلی ☺ تا اندازه ای 😐 خیر ☹ نظری ندارم ○

۴- آیا فکر می کنید این تست برای اندازه گیری مهارت های زبانی زیر مناسب می باشد؟

مهارت های زبانی

دانش عمومی زبان	general English	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
خواندن و درک	comprehension	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
نوشتن	writing	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
دستور زبان	grammer	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
دانش لغوی	vocabulary	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
هجی کردن	spelling	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○

۵- آیا نمراتی که از این تست بدست می آید عادلانه است؟

بلی ☺ تا اندازه ای 😐 خیر ☹ نظری ندارم ○

۶- آیا نتایج حاصله از این تست می تواند بخوبی نمایانگر دانش زبان آموزان باشد؟

بلی ☺ تا اندازه ای 😐 خیر ☹ نظری ندارم ○

۷- آیا دقت این تست به اندازه ای هست که بتواند بین زبان آموزان متفاوت (ضعیف، متوسط، قوی) تمایز ایجاد کند؟

بلی ☺ تا اندازه ای 😐 خیر ☹ نظری ندارم ○

۸- آیا برای پاسخ به سؤالات این آزمون، زبان آموز باید پیش از امتحان بطور جدی مطالعه کند؟

بلی ☺ تا اندازه ای 😐 خیر ☹ نظری ندارم ○

۹- این تست برای چه سطحی مناسب تر است؟ مبتدی ○ متوسط ○ پیشرفته ○ نظری ندارم ○

۱۰- درجه دشواری این تست را تا چه اندازه ارزیابی می کنید؟ آسان ○ متوسط ○ دشوار ○ نظری ندارم ○

۱۱- آیا فکر می کنید این تست برای اندازه گیری اهداف زیر مناسب می باشد؟

اهداف:

بسندگی (دانش کلی زبان)	Proficiency	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
پیشرفت (میان ترم و پایان ترم)	Achievement	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
آزمون ورودی دانشگاهها	Entrance	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
تعیین سطح	Placement	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
ضریب هوش	Intelligence	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○
استعداد زبانی	Aptitude	بلی ☺	تا اندازه ای 😐	خیر ☹	نظری ندارم ○

۱۲- کدام یک از روشهای زیر را برای تست کردن دانش زبان انگلیسی خود می پسندید؟ لطفا آنها را بترتیب علاقه خود با اعداد ۱ تا ۷ در داخل پرانتز فهرست کنید! (برای مثال، عدد ۱ نمایانگر محبوب ترین فرم تست در نظر شماست)

( ) پرکردن جای خالی Fill- the – blank ( ) تشریحی essay- type ( ) چهار جوابی multiple- choice

( ) صحیح – غلط true- false ( ) جور کردن matching ( ) کلوز – تست Cloze test

( ) سی – تست C – test

۱۳- آیا مایلید که از سی – تست برای اندازه گیری و ارزشیابی زبان انگلیسی شما استفاده شود؟

- درصورت تمایل: الف) در چه مهارتی (رجوع به سؤال ۴)؟ (ب) باچه هدفی (رجوع به سؤال ۱۱)
- در صورت عدم تمایل چرا؟

۱۴- در این قسمت خواهشمند است نظر خود را در باره هر مطلبی که فکر می کنید از قلم افتاده و یا اینکه می تواند راه گشای این تحقیق باشد مبذول فرمایید.