

# Multiple True-false (MTF) and Multiple-choice (MC) Test Formats: A Comparison between Two Versions of the Same Test Paper of Iranian NUEE

Ali Mobalegh

Department of English, Faculty of foreign languages, Esfahan University, Esfahan, Iran  
Email: ali\_mobalegh@yahoo.com

Hossein Barati

Department of English, Faculty of foreign languages, Esfahan University, Esfahan, Iran  
Email: h.barati@gmail.com

**Abstract**—This study tried to compare the Multiple-Choice (MC) item format with the Multiple True-False (MTF) item format in Iranian National University Entrance Examination (NUEE). It was desirable to find out (a) if either of the formats –MTF vs. MC– was preferred by test-takers and (b) if there was a difference between the reliability estimates of the tests with the two different formats. The study deployed a questionnaire and two tests with different formats to collect data. The analysis of the responses to the questionnaire determined that the MTF format is preferred over the traditional MC format in general by the subjects of the study. The quantitative analyses of the results of the tests show that the tests with different formats did not differ significantly in the respect of reliability coefficient.

**Index Terms**—multiple true-false (MTF) test format, cluster, multiple-choice (MC) test format

## I. INTRODUCTION

One of the fundamental concepts in language testing is *items*. Fulcher & Davidson (2007) state “it is perhaps items or tasks that first come to mind when we think of tests. This is because they are the most obvious, visible, part of a test” (p. 69). The form of a test refers to its physical appearance which depends on the form of the items comprising the test (Cohen, 1994). According to Frisbie (1992) the purpose of an item is to elicit information. Each item has two parts: one, called the stem, is for eliciting information; and the other, called the response, is for the answer to the stem. Items are classified by the figures in language testing field according to the way the stems are supposed to be responded to.

*Recognition items* require the examinees to understand the stem and then recognize and select the correct response from among several alternatives. The examinee is not asked to produce anything. The most common forms of recognition items are *true/false*, *multiple-choice (MC)*, and *matching-type* items. In true/false items, a stem is given and the examinee is to recognize whether the idea put forward in the stem is true or false (Farhady, 1985). Multiple-choice (MC) items, on the other hand, include a stem followed by a few alternatives. Only one of the alternatives is the correct response to the stem and the others are distracters. In relation to the stem of the MC items Heaton (1988) states that “the primary purpose of the stem is to present the problem clearly and concisely. ... At the same time, the stem should not contain extraneous information or irrelevant clues...” (p. 30).

Many scholars (e.g., Fulcher & Davidson, 2007; Bachman & Palmer, 1996; and Harris, 1969) believe that MC item format is the most widely used in the field of language testing. This is because of the advantages, like objectivity and high reliability, which are supposed to be inseparable with such items. Regarding the strengths of MC items, Harris (1969) maintains that “because of the highly structured nature of these items [MC items], the test writer can get directly at many of the specific skills and learnings he wishes to measure, and the examinee cannot evade difficult problems as he often can with compositions” (p. 7). Harris (1969) continues to state that “multiple-choice tests tend to have superior reliability and validity” (p. 7). McNamara (2000) points out the fact that MC tests are very versatile and can be used to assess almost all language skills. Still another advantage of MC tests refers to their ease of scoring, i.e., MC items are scored easily, rapidly, accurately and objectively by teachers, scoring machines and computers. Further, the accuracy of scoring MC tests leads to consistency of scores which in turn contributes to a higher degree of reliability (Tuckman, 1988). These advantages have caused MC tests to be used in large-scale administrations.

In spite of the above advantages, MC tests have not escaped the criticisms of the scholars. In relation to the shortcomings of MC items Heaton (1988) states that “the chief criticism of the multiple-choice item, however, is that frequently it does not lend itself to the testing of language as communication” (p. 37). Regarding another weakness of

MC items, Oller & Perkins (1978) assert, preparation of sound MC items is challenging and technically difficult. In the same way, Farhady & Shakery (2000) point out that it is commonly believed that preparing an MC item with sound and plausible alternatives is not an easy task. Some other scholars like Bush (2001), Wood (1991), Akeroyd (1982) and Coombs *et al* (1956) believe that the traditional MC tests are unable to gauge the partial knowledge due to the fact that in that format it is the matter of all or none; a test taker either answers an item or not. This way, the examiner cannot differentiate between a test taker who was doubtful between only two options and another one that did not have any idea about the correct answer. Attempts have been made to overcome such a shortcoming by proposing innovative formats to replace the traditional MC items.

One of the attempts to improve the traditional MC format has been the Multiple True-False (MTF) format. Before discussing the MTF format, which is the focus of this study, some of the formats that have tried to replace the conventional MC format are introduced in the following section.

Bush (2001) proposes an adapted new version of MC tests called “*liberal*” MC test. This innovative MC format allows test takers to choose more than one answer if they are uncertain which is the correct one. The term *liberal* is used to denote the extra dimension of choice. To appreciate the implications of selecting multiple answers, consider the scoring procedure of liberal MC items.

- (1) If an examinee knows the right answer to the question, s/he can get  $3/3=100\%$  for that question.
- (2) If the examinee is correct in thinking that the right answer is one of two options, s/he can get  $(3-1)/3 = 67\%$  for that question, compared with an equal chance of getting either 0 or 100% in a standard MC test.
- (3) If the examinee is correct in thinking that the right answer is one of three options, s/he can get  $(3 - 2)/3=33\%$  for that question, compared with having a 33% chance of getting 100% in a traditional MC test.

Another innovative trend set in MC testing was proposed by Akeroyd (1982). This is referred to as “*Dual Response*” MC test. Akeroyd (1982) proposed a scheme which is similar to liberal MC tests. It aims to assess partial knowledge explicitly and the test taker can choose more than one option. But the scoring system of *dual response* MC test is different. For an item with 4 alternatives, 1 mark is awarded for selecting the correct answer, 0.5 for selecting two answers including the correct one, 0.25 for an omission of a distracter and 0 for anything else. Bush (2001) believes that this method is inferior to his *liberal* MC test for three reasons: “(1) it appears to be less rational; (2) it only rewards examinees who successfully select one or two answers (not three); and (3) it results in inflated marks overall, even though it does not reward examinees as much as our method for selecting two answers that include the correct one” (p. 162).

Another method which asserts to have some advantages over the traditional MC tests is the “*Elimination Procedure*” proposed by Coombs *et al.* (1956). This requires examinees to select the answers which they believe are *wrong*, rather than those they believe are or may be right. The tests items are marked +1 when the wrong answer is correctly eliminated, but  $n-1$  is for every right answer eliminated, while “ $n$ ” is the number of answers per question.

Finally, the Multiple True-False format (MTF) is another strive to enhance the MC format. The origin of the format is not clear in the literature but the most influencing propagator of it are Frisbie (1992), Frisbie, D. A., & Becker, D. F. (1991), Frisbie, D. A., & Druva, C. A. (1986), and Frisbie, D. A., & Sweeney, D. C. (1982). It is worth mentioning that no studies on this format have been done in recent years. Since this format is the focus of this study, it is explained thoroughly in subsequent section.

## II. BACKGROUND

### A. Definition of MTF Items

MTF tests are classified as *recognition* type tests which have been presented only in written mode. Frisbie (1992) defines the MTF item as:

...one of several pieces of a cluster that looks like a multiple-choice item. The cluster has a stem that is an incomplete statement, and it has two or more options that independently complete the stem. The examinee must respond *true* or *false* to each statement formed by the stem and an option (p. 21).

Four points are very important in the definition of the MTF item given by Frisbie (1992): 1. One MTF cluster consists of the stem plus the following options. 2. Each MTF *cluster* contains as many *items* as the options that follow the stem. 3. The stem of an MTF item should be an *incomplete* statement, not a question. 4. Considering the preceding stem, the examinee is supposed to decide if each of the options is *true* or *false*, rather than select only one option as the correct answer, as he/she does in an MC test. Table1 displays 2 MTF cluster; they are not for testing language ability.

TABLE 1.  
TWO MTF CLUSTERS

*** THE SURVEY METHOD OF RESEARCH USUALLY INVOLVES ____.
___ 1. EXPERIMENTAL TECHNIQUES
___ 2. NON-MANIPULATIVE TECHNIQUES
___ 3. INVASION OF PRIVACY
___ 4. CORRELATION METHODS
*** THE SET OF SCORES (4, 5, 5, 7, 8, 9, 9, 15) HAS ____.
___ 5. A MEDIAN OF 7.5
___ 6. A SYMMETRIC SHAPE
___ 7. MORE THAN ONE MODE
___ 8. A LARGER MEAN THAN MEDIAN

The two clusters shown in table 1 are about principles of research and mathematics respectively. Two things are important to note:

1. As shown in table 1, the beginning of each cluster is identified with three asterisks (\*\*\*) and the stems are not numbered; the options, rather than the stems, are numbered. And it is important to point out that the options of the clusters are numbered consecutively; the same way of numbering would continue to the next clusters. The same layout of the MTF item format has been followed by the figures in the field since many years ago (e.g., Dudley, 2006, 2004; Brown & Hudson, 2002; Frisbie, 1992; Frisbie & Sweeny, 1982; Cronbach, 1939).

2. The second important point about table 1 is that each MTF cluster, which looks like one MC item, contains 4 items; so by converting each MC item to an MTF cluster, 4 MTF items are formed.

About the MTF item format Frisbie (1992) states:

Each MTF cluster may have any number of items true, and the correctness of any one item is judged independently of the correctness of other items within the same cluster. ...unlike an MC item stem, the MTF cluster stem cannot be a question, and it should not ask the respondent to compare options in any way (p. 22).

So in one MTF cluster one, two, three or all of the options (items) could be true. From Frisbie's (1992) comment, it can be concluded that all the items in the same cluster are not about the same point necessarily; each option (item) in the cluster should be looked at as an independent question from other options in the same cluster. A cluster, then, may deal with a single idea or with multiple related ideas. Both clusters in table 1 deal with multiple related ideas.

An important aspect of the MTF item format is the way in which this format is scored. The following section explains briefly the ways that figures in the field have used before.

#### B. Ways of Scoring MTF Items

One of the areas which is not crystal clear in the literature is the way MTF items should be scored; none of the sources available clarified the subject thoroughly. However, Dudley (2006) and Frisbie & Druva (1986) briefly explain that there are two scoring methods: *cluster scoring* and *primary scoring*. In the former method, items are scored in a similar way to the way traditional MC items are scored. That is, each cluster receives a score based on the number of options correctly identified, true or false, for its stem. In the latter method, the primary scoring, the test is scored based on the total number of correct responses, irrespective of the clusters.

This study particularly focuses on the MTF format and the comparison between the MC and the MTF format. The advantages that are talked about in the literature are discussed in the next section.

#### C. Advantages of the MTF Item Format

Many studies have been conducted to compare the MTF format with other formats that are popular. Almost all of these studies (Hill & woods, 1974; Albanese & Sabers, 1988; Gross, 1978; Frisbie & Sweeny, 1982; Frisbie & Druva, 1986; Kreiter & Frisbie, 1989 and Dudley, 2004, 2006) agree that:

1) The number of main responses elicited in a particular amount of time is more in MTF format than in other similar formats.

2) The reliability estimates of items in the MTF format is equal to or higher than those of the items in other formats like MC items.

Frisbie & Sweeney (1982) report that "the expected advantage in terms of test reliability was documented by Albanese et al. (1977), Kelly (1979), and Mendelson et al. (1980)" (p. 29).

In line with the advantages of the MTF format Kreiter & Frisbie (1989) point out:

The MTF format elicits four responses per item stem, whereas the MC item permits only one response. If the examinee answers the MC item correctly, it is difficult to make a meaningful inference regarding the individual's knowledge about the unmarked responses. However, because a response is made to each of the four items in the MTF cluster, the altered format [MTF format] appears to yield more information than the conventional MC item (p. 208).

The advantages explained above have made the MTF format an important item format in the field of testing. According to Dudley (2006) "research on the validity of using of the MTF format as a testing tool in general education has been carried out since the 1940s, and its value has been well documented in both early and recent testing literature" (p. 199). In the field of second language testing, however, caution has thus far been advised in using the MTF format because its application to second language testing has not undergone rigorous research (Brown and Hudson, 2002).

### III. SIGNIFICANCE OF THE STUDY

In Iran, almost all of the important high-stakes language tests, like the entrance exam of universities, are administered in the MC format. Very few people are acquainted with the MTF format which can be a strong rival for the MC format. Through this study, the researcher aims to discover if the English language items in the MTF format can exceed their counterparts in the MC format in national university entrance examination, NUEE, in Iran. The researcher was supposed to find out about students' perception of the MTF format compared to their perception of the MC format. Also, it is desirable to compare the reliability estimates of the same test with the two different formats, MC and MTF.

### IV. RESEARCH QUESTIONS

The study focused on two research questions:

1. Is either of the formats –MTF vs. MC– preferred by test-takers?
2. Is there a difference between the reliability estimates of the tests with the two different formats?

### V. METHODOLOGY

#### A. Pilot Study

At the beginning, MTF items were constructed out of the MC items of the English part of the NUEE administered in 1386. So, MTF clusters were constructed out of 20 MC items. Since the tests meant to be used in the main study had not been piloted before, the researcher piloted the tests to make sure no problems would crop up in the main study. This pilot study was to ensure the researcher of the appropriate timing, and test administration procedures. The 48 participants in the pilot study were selected from two typical public high schools by chance. All the participants were male students aged 18 to 19. 24 participants in the pilot study were given the MC test and 24 participants in the pilot study took the MTF test. The participants comments on the MTF items compared with the MC items were recorded to be used in the closed questionnaire used in the main study. The pilot study had some implications for the main study: 1. It was found that test takers had some problems understanding test instructions, therefore it was decided to change the instructions into Persian. 2. 20 students of the group who took the MTF test (20 clusters) finished the test within 30 minutes, so it was decided that 30 minutes would be sufficient for the subjects in the main study to take the MTF test.

3. None of the students in the groups who took the MC test (20 items) had any problems with the test and 22 of them finished the test in 15 minutes. Again it was found out that 15 minutes would be enough for the MC test to be taken.

#### B. Main Study

Investigating if the MTF format can be a good substitute for the MC format in general and in the NUEE in particular gave rise to the present study. The focus of the research is whether prospective test takers prefer the MTF format over the traditional MC format and if the MTF format enjoys a better reliability coefficient.

##### a. Participants

For carrying out the main study 200 students of the pre-university level were selected. All the participants were 18- to 19-year old boys. The participants were selected through cluster random selection. Two districts, 1 and 5 were selected. In each district two schools and in each school two classes were selected.

TABLE 2.  
NUMBER OF THE PARTICIPANTS AND THEIR SPECIFICATIONS IN THE MAIN STUDY

DISTRICT	PRE-UNIVERSITY SCHOOL	CLASS	N	GENDER	AGE
5	SADAT	A	25	MALE	18-19
		B	25	MALE	
	GHADIR	A	25	MALE	18-19
		B	25	MALE	
1	SHEIKHBAHAI	A	25	MALE	18-19
		B	25	MALE	
	NILFOROUSHZADEH	A	25	MALE	18-19
		B	25	MALE	

##### b. Instruments

As mentioned above this study aims to examine if converting the format of NUEE from MC to MTF can make the test more satisfying for the test takers. To do so, two tests with different formats (i.e., MC and MTF) were developed. In addition, one questionnaire was constructed using the comments given by participants in the pilot study. The instruments are explained in detail below.

##### 1. The MC Test

One of the tests used in the study was in MC format. The test was one of the NUEE papers administered in 1386. The English language test in the first booklet given to the applicants for the universities was selected. It is also important to know that the last part of the test, which consisted of 5 reading comprehension questions, was omitted from the study because reading comprehension questions were not included in the MTF test. The MC test used in the study consisted of 20 items; 5 items on English grammar, 10 items on English vocabulary, and 5 cloze items. Except omitting the

reading comprehension questions, the MC test was not changed anymore; in the pilot study no problems were encountered so no changes were necessary (see appendix1).

## 2. The MTF Test

This test was developed by the researcher and was the converted MTF format of the MC test discussed above. The reading comprehension questions of the MC test were not converted and were not used in the study. It was done due to the fact that it was not possible to consider more than one correct answer for each reading comprehension question; so converting them to MTF clusters was meaningless. After converting the MC items, the MTF items were revised by 2 university professors and 3 postgraduates of applied linguistics and 6 high school English teachers who had more than 7 years of experience in English teaching. Their comments on test rubrics and some items were taken into account and a few changes were made in some items and instructions. The MTF test developed by the researcher had 80 items (20 items on English grammar, 40 items on English vocabulary, and 20 cloze items) that corresponded with the 20 MC items of the NUEE test. Because the tests in the MTF format consist of clusters and each cluster includes 4 *true/false* items the number of items are multiplied by four when the MC tests are converted to MTF tests; each MC item is converted to one MTF cluster (see appendix2).

## 3. The Questionnaire

The other instrument used to collect data in the present study was a closed Likert scale 4-degree questionnaire (see appendix3) to be answered by the subjects in the study. This questionnaire had 10 questions based on the comments made by the participants in the pilot study and it was in Persian so that participants would not have difficulty understanding the sentences. The respondents were allowed to ask any questions for clarification if they needed.

## c. Procedure

The study was carried out in each school separately. The participants were asked to congregate in the auditorium of the school. The seats were arranged in 4 columns. The participants in the first and third columns took the MTF test and those in the second and fourth columns took the MC test.

On the first page of each test some points were explained. Those points included the time allowed for completing the test (15 minutes for the MC test and 30 minutes for the MTF test), the penalty for each wrong answer and that they should use the answer sheet available to them for answering the test.

As discussed above, the instructions of the MTF test was in Persian but the instructions of the MC test was in English, as it is in the NUEE sessions.

The students were asked to start the test at a specific time. Attempt was made to make the test situation exactly like NUEE sessions. Dictionaries were not allowed and the participants were prohibited from talking with each other. Also, no questions were answered to keep up with the NUEE typical situation.

After 15 minutes all the students in the second and fourth columns, who took the MC test, were asked to hand in their answer sheets and after 30 minutes students in the first and third columns, who took the MTF test, were asked to hand in their answer sheets.

Because the participants' teachers agreed to consider the results of the test as a determining factor for the participants' final mark, the participants in the study were presumably motivated enough to take the test seriously. It is also important to know that this study was carried out right after English language final exam in each school, so test takers were prepared as they are for NUEE.

The participants who took the MTF test were familiar with the traditional MC format of NUEE. They had been studying at high school for 4 years and had experienced taking many MC tests. So they were assumed as qualified for comparing the two formats of MC and MTF. So following the exam session, those participants who took the MTF test completed the questionnaire on the new test, the MTF test.

## d. Scoring Procedures

For scoring procedures the MC test items received +1 for each correct response, -1/3 for each wrong answer and the questions left unanswered were given 0; this is the procedure implemented in NUEE.

Because each cluster in the MTF test corresponds to one MC item, attempt was made to apply a method comparable to that used for the MC test. So it was decided to draw on the *cluster scoring* method for scoring the MTF tests. One participant who answered all the items in one cluster correctly should get the same score as the participant who answered one MC item correctly. Each correctly answered MTF item received +1/4, so if a participant answered all the four items of a cluster correctly he received +1. And each wrong response to MTF items received -1/12, so if a participant answered all the four items of a cluster incorrectly he received -1/3. Also, unanswered MTF items were given 0.

# VI. DATA ANALYSIS AND RESULTS

All the gathered data were put into analytical techniques to address the research questions. In order to sum up the participants' opinion about the MTF format and to come up with a general conclusion for the first research question, *Mann-Whitney* test analysis was run on the *average* of the answers to the questions of the questionnaire. The options of the questions in the questionnaire were given different values; *completely agree*: 4, *agree*: 3, *disagree*: 2, *completely disagree*: 1; the average value of the options is 2.5,  $(4+3+2+1)/4=2.5$ . This analysis is presented in table 3 below.

TABLE 3.  
RESULTS OF MANN-WHITNEY ANALYSIS ON THE AVERAGE OF RESPONSES TO THE QUESTIONNAIRE  
Test Statistics(a)

	AVG1
Mann-Whitney U	4000.000
Wilcoxon W	9050.000
Z	-2.612
Asymp. Sig. (2-tailed)	.009

a Grouping Variable: VAR00001  
Statistics  
AVG1

<b>N</b>	<b>Valid</b>	<b>100</b>
<b>Mean</b>		<b>2.7267</b>

As it is shown above Sig. value is .009. The Sig. value is less than the critical value which is .05. So the two difference between the formats is significant. Because the *mean* value is 2.72 it can be concluded that the MTF format is generally preferred over the MC format by the participants; 2.72 is greater than 2.5 which is the average of the values of the responses to the questions in the questionnaire.

The reliability analysis of Cronbach's Alpha on the results of the MC and MTF tests proved that the reliability estimates of the tests with different formats were not very different; the reliability value of the MTF test was 0.78 and the reliability value of the MC test was 0.79. These results are unexpectedly very different from the previous research (e.g., Dudley, 2004; Frisbie & Sweeny, 1982; etc).

## VII. DISCUSSION

It should be pointed out that some aspects of the MC format are preferred over those of the MTF format. They are listed below.

1. In the fifth question of the questionnaire, the participants were asked if the *cloze test* of the MTF exam was boring, tiring and time consuming. On the basis of the data obtained from the responses, 68 percent of the participants voted that the MTF test was boring, tiring and time consuming in the cloze part. This can be considered one of the shortcomings or limitations of this new format. No previous research has worked on the effectiveness of the MTF format in cloze testing procedure.

2. In the MTF tests the test taker has to judge the same incomplete sentence (stem) for as many times as the number of the following options (items). In the sixth question of the questionnaire the participants were asked if it was really awkward to do that. Based on the responses to question 6 of the questionnaire, 62 percent of the participants agreed that they were not comfortable reading the stem of each cluster for four times.

3. The ninth question of the questionnaire asked the participants if the MTF test was more confusing and more difficult to understand than the MC format. On the basis of the responses to question 9, 71 percent of the participants believed that the MTF format was confusing. One factor that could have caused it is the fact that test takers had never encountered the MTF format up to then. It should be reminded that any new unconventional format would have been as confusing in the first administration. This confusion might have stemmed from other factors, too. Anyway, the reason is not related to the present study.

All in all, the MTF format can be looked at as a proper substitute for the traditional MC format in the NUEE in particular and in any high-stakes achievement tests in general. This study has proved that the MTF format enjoys a better face validity and acceptance from the test taker perspective. This claim is based on the definition given for face validity by McNamara (2000): "Face validity is the extent to which a test meets the expectations of those involved in its use, e.g. administrators, teachers, candidates and test score users" (p. 133).

## VIII. LIMITATION OF THE STUDY

Throughout the study, the researcher encountered some problems and limitations which are listed below from the most serious limitation to the least serious one:

1. The MTF format could not be used for the *reading comprehension* part of the English language test in *konkoor*. In fact, all reading comprehension questions can have only one correct answer and using MTF clusters is meaningless for them.

2. Another obstacle that the researcher had to struggle with was motivating the students to prepare themselves for the test and take the test seriously. Of course, the study was carried out after the participants' final English exam, so the participants were prepared. To motivate the students to take the test seriously, their teachers agreed to consider the result of the tests as one part of the participants' final exam.

3. Because of the rules of the educational ministry, the researcher was not allowed to go to girls' high schools to gather data from girls; this is the reason why the participants in the study were only boys.

## IX. SUGGESTIONS FOR FURTHER RESEARCH

The findings of the present study together with the limitations observed above provide some clues for further research studies. The following suggestions may be helpful:

1. In the present study the concurrent validity of the MTF test was not checked against a standardized test like TOEFL. It is important to investigate the extent to which the tests with MTF format are concurrently valid to such tests.
2. This study can be carried out including female participants and also the same study can be done in a nation-wide scale.

## X. CONCLUSION

In conclusion, this study has instigated the way of replacing the MC format with the MTF format. It has been proved that the replacement has many precious advantages like a better face validity. This change calls for more research and other aspects of it should be investigated in further studies.

## APPENDIX 1: THE MC TEST OF NUEE HELD IN 1386

### Part A: Grammar

**Choose the one word or phrase that best completes the sentence.**

1. I turned off the TV because I didn't find the film \_\_\_\_\_.  
1) amusingly 2) amusing 3) amused 4) amuses
2. Mary is taking driving lessons \_\_\_\_\_ she can easily pass the test.  
1) when 2) whether 3) because 4) so that
3. It is \_\_\_\_\_ late for us to go to the movie tonight.  
1) very 2) so as 3) too 4) enough
4. All the food \_\_\_\_\_ before we got to the picnic.  
1) has eaten 2) had eaten 3) had been eaten 4) has been eaten
5. Ann's birthday was on the 5<sup>th</sup>, and today is the 8<sup>th</sup>. Her birthday card \_\_\_\_\_ a week ago.  
1) should be sent 2) should have sent  
3) should be sending 4) should have been sent

### Part B: Vocabulary

**Choose the one word or phrase that best completes the sentence.**

6. A thin flat piece of paper, metal, plastic, glass, etc. is called \_\_\_\_\_.  
1) sheet 2) core 3) layer 4) crust
7. The examination is going to be \_\_\_\_\_ at 10:30. Try not to come late.  
1) made 2) held 3) performed 4) occurred
8. Jane can get the job done. I have full \_\_\_\_\_ in her abilities.  
1) confidence 2) exploration 3) endeavor 4) training
9. Could you stop talking please? You are \_\_\_\_\_ me from my work.  
1) recalling 2) influencing 3) denying 4) distracting
10. My son does not wear these clothes. He says they are not \_\_\_\_\_ these days.  
1) forbidden 2) emotional 3) fashionable 4) superior
11. There is too little \_\_\_\_\_ about the present state of the German economy.  
1) certainty 2) attention 3) emergency 4) presentation
12. Were you \_\_\_\_\_ to move away from home by your parents?  
1) located 2) predicted 3) transferred 4) encouraged
13. I'm afraid there is no mechanism for dealing with your complaint. "Mechanism" means:  
1) purpose 2) method 3) function 4) occasion
14. I was sick and had to stay home for two weeks. Now I think it is hard to \_\_\_\_\_ up with the class.  
1) give 2) keep 3) pass 4) take
15. From his \_\_\_\_\_ expression I realized he was sad but not happy.  
1) reasonable 2) natural 3) facial 4) private

### Part C: Cloze Test

**Read the passage and decide which choice (1), (2), (3), or (4) best fits each space.**

Language may have started in one place and spread from that place. Or it may, 16) \_\_\_\_\_, have started in several places. In any event, if the people of the world had been 17) \_\_\_\_\_ in touch with one another all through their history, we might now have had only one language. 18) \_\_\_\_\_ mountains and seas have 19) \_\_\_\_\_ peoples apart. Today there are more than 1000 languages. Some 20) \_\_\_\_\_ by many millions of people, others by only a few thousand.

16. 1) so far 2) instead 3) thus 4) besides
17. 1) closely 2) briefly 3) silently 4) softly
18. 1) But 2) Therefore 3) Though 4) While
19. 1) removed 2) made 3) kept 4) looked
20. 1) spoken 2) be spoken 3) were spoken 4) are spoken

## APPENDIX 2: THE MTF TEST (THE CONVERTED TEST)

در هر مستطیل یک جمله ناقص وجود دارد و هر گزینه یک سوال محسوب می شود ( هر مستطیل حاوی چهار سوال است). جمله های ناقص را بخوانید و در پاسخنامه درستی یا نادرستی هریک از گزینه ها را به صورت جزا مشخص کنید.  
 گزینه صحیح: T  
 گزینه غلط: F

*There is an incomplete sentence in each box. Decide which options correctly complete the sentence (T) and which ones cannot correctly complete the sentence (F).*

## Part A: Grammar

I turned off the TV because \_\_\_\_\_.

- 1) it wasn't amusing T / F  
 2) I wasn't amused T / F  
 3) it did not amuse me T / F  
 4) I wasn't amusing T / F

Mary is taking driving lessons \_\_\_\_\_.

- 5) when she can pass the test T / F  
 6) whether it is easy or not T / F  
 7) because she knows how to drive T / F  
 8) so that she can easily pass the test T / F

It is \_\_\_\_\_ late for us to go to the movies tonight.

- 9) so T / F 10) really T / F 11) too T / F 12) enough T / F

All the food \_\_\_\_\_ before we got to the picnic.

- 13) had been eaten T / F  
 14) was finished T / F  
 15) had eaten T / F  
 16) has eaten T / F

Ann's birthday was on the 5th, and today is the 8th. Her birthday card \_\_\_\_\_ a week ago.

- 17) should be sent T / F 18) was sent T / F  
 19) had been sent T / F 20) should have been sent T / F

## Part B: Vocabulary

A thin flat piece of paper, metal, plastic, glass, etc. is called \_\_\_\_\_.

- 21) sheet T/F 22) core T/F 23) page T/F 24) crust T/F

The examination is going to \_\_\_\_\_ at 10:30. Try not to come late.

- 25) be given T/F 26) be held T/F 27) begin T/F 28) take T/F

Jane can get the job done. I have \_\_\_\_\_ in her abilities.

- 29) no exploration T/F 30) endeavor T/F  
 31) no doubt T/F 32) confidence T/F

Could you stop talking please? You are \_\_\_\_\_.

- 33) disturbing me T/F 34) interrupting the speech T/F  
 35) distracting my attention T/F 36) recalling me T/F

My son does not wear these clothes. He says they are not \_\_\_\_\_ these days.

- 37) forbidden T/F 38) emotional T/F  
 39) fashionable T/F 40) popular T/F

There is too little \_\_\_\_\_ about the present state of the German economy.

- 41) certainty T/F 42) presentation T/F 43) emergency T/F 44) hope T/F

Were you \_\_\_\_\_ to move away from home by your parents?

- 45) permitted T/F 46) allowed T/F 47) transferred T/F 48) encouraged T/F

I'm afraid there is no mechanism for dealing with your complaint.

"Mechanism" means:

- 49) way T/F 50) method T/F 51) function T/F 52) occasion T/F

I was sick and had to stay home for two weeks. Now I think it is hard to \_\_\_\_\_ up with the class.

- 53) give T/F 54) keep T/F 55) catch T/F 56) take T/F

From his \_\_\_\_\_ expression I realized he was sad but not happy.

- 57) reasonable T/F 58) natural T/F 59) facial T/F 60) obvious T/F

## Part C: Cloze Test



متن زیر را به دقت بخوانید. برای هر جای خالی در متن چهار گزینه وجود دارد. در پاسخنامه برای هر گزینه که میتواند جای خالی مربوط به خود را کامل کند T را علامت بزنید و برای هر گزینه که نمیتواند عبارت مناسبی برای جای خالی مربوطه باشد F را علامت بزنید.

Read the text below and for each gap decide which options are appropriate and true (T) and which options are inappropriate and false (F).

Language may have started in one place and spread from that place. Or it may, (61-64) \_\_\_\_\_, have started in several places. In any event, if the people of the world had been (65-68) \_\_\_\_\_ in touch with one another all through their history, we might now have had only one language. (69-72) \_\_\_\_\_ mountains and seas have (73-76) \_\_\_\_\_ peoples apart. Today there are more than 1000 languages. Some (77-80) \_\_\_\_\_ by many millions of people, others by only a few thousands.

61) so far T/F	62) instead T/F	63) also T/F	64) besides T/F
65) closely T/F	66) quietly T/F	67) silently T/F	68) well T/F
69) But T/F	70) Therefore T/F	71) While T/F	72) However T/F
73) removed T/F	74) called T/F	75) kept T/F	76) held T/F
77) spoken T/F	78) are spoken T/F	79) be spoken T/F	80) were spoken T/F

### APPENDIX 3: THE QUESTIONNAIRE GIVEN TO THE STUDENTS PARTICIPATING IN THE STUDY

لطفاً سوالات زیر را به دقت بخوانید و گزینه مورد نظر خود را انتخاب کنید. از همکاری شما کمال تشکر را داریم.

Please, read the questions and choose one of the options for each question.

- I think the MTF test can cover more parts of the text book than the MC test.  
a) completely agree b) agree c) disagree d) completely disagree  
1. به نظر من این نوع امتحان نسبت به سؤالات چهار گزینه ای مطالب بیشتری از کتاب را در یک سوال می سنجید.  
(الف) کاملاً موافقم (ب) موافقم (ج) مخالفم (د) کاملاً مخالفم
- I think the MTF test let me show my abilities better.  
a) completely agree b) agree c) disagree d) completely disagree  
2. فکر می کنم این نوع سوال کمک کرد تا تواناییهای خود را در زبان انگلیسی بهتر نشان دهم.  
(الف) کاملاً موافقم (ب) موافقم (ج) مخالفم (د) کاملاً مخالفم
- The MTF test made me think more deeply than the MC test.  
a) completely agree b) agree c) disagree d) completely disagree  
3. این سبک سوال نسبت به سؤالات چهار گزینه ای من را وادار به فکر کردن بیشتری می کرد.  
(الف) کاملاً موافقم (ب) موافقم (ج) مخالفم (د) کاملاً مخالفم
- In the vocabulary part of the MTF test I was given a point for each option the meaning of which I knew and was able to say whether it fit the blank or not, which was really fair.  
a) completely agree b) agree c) disagree d) completely disagree  
4. در بخش دوم (Vocabulary) چون معنای لغات داخل هر گزینه را که میدانستم برایم تکرار مثبت داشت بسیار خوب بود.  
(الف) کاملاً موافقم (ب) موافقم (ج) مخالفم (د) کاملاً مخالفم
- The cloze test of the MTF test was *not* boring or time consuming.  
a) completely agree b) agree c) disagree d) completely disagree  
5. برای بخش سوم (Cloze Test) این نوع سوال هرگز خسته کننده و وقت گیر نبود.  
(الف) کاملاً موافقم (ب) موافقم (ج) مخالفم (د) کاملاً مخالفم
- That I had to read each stem four times, once for each option, was really trouble making.  
a) completely agree b) agree c) disagree d) completely disagree  
6. اینکه مجبور بودم هر جمله ناعلم را چهار بار بصورت جداگانه بخوانم مشکل ساز نبود.  
(الف) کاملاً موافقم (ب) موافقم (ج) مخالفم (د) کاملاً مخالفم
- The MTF test raised new questions for me and made me notice important gaps in my knowledge.  
a) completely agree b) agree c) disagree d) completely disagree  
7. امتحان چنان بودند که برای من ایجاد سوال های جدید می کرد و من را متوجه نکاتی که در آنها نیاز به یادگیری بیشتر داشتم میکرد.  
(الف) کاملاً موافقم (ب) موافقم (ج) مخالفم (د) کاملاً مخالفم
- In the traditional MC tests, knowing the answer- the correct option- was enough but in the MTF test I needed to have a better command to get the whole points.  
a) completely agree b) agree c) disagree d) completely disagree  
8. در سوال های چهار گزینه ای اگر جواب درست را می دانستم کافی بود ولی در این سوال ها نیاز به تسلط بیشتری به مطالب کتاب داشتم.  
(الف) کاملاً موافقم (ب) موافقم (ج) مخالفم (د) کاملاً مخالفم
- The MTF test was *not* confusing in general.  
a) completely agree b) agree c) disagree d) completely disagree

9. این نوع سؤال در مجموع اصلاً مبهم و گیج کننده نبود.  
الف) کاملاً موافقم ب) موافقم ج) مخالفم د) کاملاً مخالفم
10. I think that the MTF format was quite different from the regular true-false format.  
a) completely agree b) agree c) disagree d) completely disagree
10. بنظر من این نوع سؤال کاملاً با سوالات درست/نادرست (T/F) معمولی متفاوت است.  
الف) کاملاً موافقم ب) موافقم ج) مخالفم د) کاملاً مخالفم

## REFERENCES

- [1] Akeroyd, F. M. (1982). Progress in multiple-choice scoring methods. *Journal of Further and Higher Education*, 6, 87–105.
- [2] Albanese, M. A., Sabers, D. L. (1988). Multiple true-false items: A study of interitem correlations, scoring alternatives, and reliability estimation. *Journal of Educational Measurement*, 25, 111-23.
- [3] Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- [4] Brown, L.D., & Hudson, T. (2002). *Criterion-referenced language testing*. New York, NY: Cambridge University Press.
- [5] Bush, M. (2001). A multiple choice test that rewards partial knowledge. *Journal of Further and Higher education*, 25 (2), 157-64.
- [6] Cohen, A. D. (1994). *Assessing language ability in the classroom* (2<sup>nd</sup> ed.). New York: Heinle and Heinle.
- [7] Coombs, C. H., Milholland, J. E., & Womer, F. B. (1956). The Assessment of partial knowledge. *Educational and Psychological Measurement*, 16, 13-37.
- [8] Cronbach, L. J. (1939). Note on the multiple true false exercise. *Journal of Educational Psychology*, 30, 628-31.
- [9] Dudley, A. (2004). *The validity of multiple true-false test format in second language testing*. Unpublished doctoral dissertation. Temple University.
- [10] Dudley, A. (2006). Multiple dichotomous-scored items in second language testing: Investigating the multiple true-false item type under norm-referenced conditions. *Language Testing*, 23 (2), 198-228.
- [11] Farhady, H. (1985). Forms of language tests. *Roshd Foreign Language Teaching Journal*, 1 (3), 76-92.
- [12] Farhady, H., & Shakery, S. (2000). Number of options and economy of multiple-choice tests. *Roshd Foreign Language Teaching Journal*, 14 (57).
- [13] Frisbie, D. A. (1992). The multiple true-false item format: A status review. *Educational Measurement: Issues and Practice*, 11, 21-6.
- [14] Frisbie, D. A., & Becker, D. F. (1991). An analysis of textbook advice about true-false tests. *Applied Measurement in Education*, 4 (1), 67-83.
- [15] Frisbie, D. A., & Druva, C. A. (1986). Estimating the reliability of multiple true-false tests. *Journal of Educational Measurement*, 23, 99-105.
- [16] Frisbie, D. A., & Sweeney, D. C. (1982). The relative merits of multiple true-false tests. *Journal of Educational Measurement*, 19, 29-35.
- [17] Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York, NY: Routledge.
- [18] Gross, L. J. (1978). Considerations in scoring multiple true false tests. *Health Professions Education Bulletin*, 7, 26-30.
- [19] Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.
- [20] Heaton, J. B. (1988). *Writing English language tests*. New York: Longman.
- [21] Hill, G. C. & Woods, G. T. (1974). Multiple true-false questions. *Education in Chemistry*, 11, 86-87.
- [22] Kreiter, C. D., & Frisbie, D. A. (1989). Effectiveness of multiple true-false items. *Applied Measurement in Education*, 2, 207-16.
- [23] McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- [24] Oller, J. W., Jr. & Perkins, K. (Eds.), (1980). *Research in language testing*. Rowley, MA: Newbury House.
- [25] Suen, H. K. (1990). *Principles of test theories*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- [26] Tuckman, B. W. (1988). *Testing for teachers*. New York: Harcourt Brace Jovanovich.
- [27] Wood, R. (1991). *Assessment and testing: A survey of research*. Cambridge: Cambridge University Press.



**Ali Mobalegh** (b. 1982, Isfahan, Iran) is currently Master from the University of Isfahan, Isfahan, Iran. He received his M.A. in TEFL at the University of Isfahan (2008) following the completion of his B.A. in English Literature from University of Isfahan, Iran (2005). His main research areas of interest are: English Language Testing, Language Teaching Methodology. He has been working as an EFL instructor since 2000.



**Hossein Barati** works in the English department, University of Isfahan, where he is assistant Professor of Applied Linguistics. He has a PhD in 'Language Testing' from University of Bristol and has been involved in research in language testing, programme evaluation, reading strategies, and classroom discourse. Before becoming a university lecturer in 1992, he worked as a language teacher at Ministry of Education, Isfahan, Iran. He has published in areas of language testing and assessment, and language programme evaluation. He is a member of a research project on *Developing a new model of teaching to adult English non-majors at University of Isfahan*, funded by University of Isfahan.

He has also co-authored some publications on DIF in Iranian National University Entrance Exam (INUEE), 2007; Linguistic constraints and language teaching, 2008; and Cultural differences in EFL performance on cloze tests, 2010. Further a text-book *A step forward in English for intermediate EFL learners* (2010) is among his latest publications.

Hossein is currently working on 'New plans for Teaching English to Iranian adult non-majors', 'Teaching English to Iranian young learners, and 'The consequential validity of high stakes tests in the Iranian context'.