The Effects of Test Formats on the Performances of Iranian EFL Students

Zahra Shahivand

Department of English Language, Andimeshk Branch, Islamic Azad University, Andimeshk, Iran

Abbas Paziresh Department of Translation Studies, Payamnoor University, Iran

Alireza Raeeszadeh

Department of English language, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

Abstract—This study was an attempt to investigate whether different test formats had any significant effect on the performances of Iranian language students. To fulfill the purpose of this study, 160 male and female language students within the age range of 20 to 34 years studying at three branches of Islamic Azad university-Ahvaz, Dezful, Andimeshk at three different levels of upper-intermediate, intermediate and low were selected based on their performance on Nelson Test (Fowler & Coe, 1976) and assigned to three groups in terms of their proficiency level. Each group of each level of language abilities was given a test consisting of four different formats-MC, T/F, Cloze-Test and C-Test. The results were analyzed based on descriptive statistics and deductive statistics of ANOVA and Scheffe Test. The results proved that the testees of upper-intermediate level had a poorer performance in MC items than they had in other test formats, while the testees of intermediate and low level performed much better in MC items; of course they had a poor performance in Cloze-Test. Due to their limited knowledge of language and limited mental abilities and strategies, testees of intermediate and low level performed very poorly in integrative tests as these test formats take high mental activities. The results proved the use of multi-method theory in testing language abilities, as it will is assumed a suitable measure in order to make judgments about the test construct validity and the testees' performance.

Index Terms-test formats, testees' performances, integrative test items, discrete test items

I. INTRODUCTION

Teaching and testing have an interdependent relationship, because teaching process needs to test different taught materials such as acquired knowledge during teaching process, to examine test administration methods, to evaluate to some extent the success and the failure of teaching goals. Also, testing depends on teaching, in as much as there are some to teach, there are some to test, in order to receive feedback and to make sound decisions to achieve a good teaching and learning process in future (Farhady, Ja'farpur & Birjandi, 2004; Hughes, 2003). According to Brown (1995), language evaluation is a pre-determined procedure to gather information to consider them to make better and effective teaching-learning process. Also, it is needed to measure the learners' competence. Language testing as a means to measure the learners' competence and their language knowledge plays an important function in learning process.

According to Cronbach and Meehle (1955), construct is referred to the assumed characteristics of individuals, supposed to be returned to the test performance. In validating a test, the qualities upon which we give some statements in test interpretation are of utmost importance. Also, Brown (2000) claimed that construct is "an attribute, proficiency, ability, or skill that happens in the human brain and is defined by established theories, e.g. overall English language proficiency" (p. 9). Downing (2003) claims that validity is the fair, unbiased, and systematic method in data gathering from various sources and He adds that "assessments are not valid or invalid; the scores or outcomes of assessments have more or less evidence to support or to refute a specific interpretation(p. 830). In another view, Clapham (2000) mentions that validity is referred to test scores, not the test itself. He adds that the test validation is an observational measurement of test scores, test results, and test interpretations. Cohen (2001) believes that, validity is "whether the test actually measures what is a purport to measure" (p. 525). So, Cohen (2001) suggests a definition of construct validity as such, " this form of validity refers to the degree to which scores on a measure permit inferences about underlying traits, and it examines whether the test is a true reflection of the theory of the trait being measured" (p. 526). Also, Cronbach and Meehle (1955) suggested that whenever the correlation between two tests is low, the construct validity is also low.

Evaluation through various methods gives a comprehensive and thorough view of what the learners have achieved during the course of study. The use of multi-methods in test presentation may be rationale to measure and to investigate

the construct validity of the tests. Different perspectives of test facets and validity in general, and construct validity particularly highlighted by many different scholars during the history of language teaching and testing. Baghaei (2009) believed that when a testee receives two test formats from the same tested content, s/he should have the same test results and performance. According to Bachman and Palmer (2010), the method used to elicit testees` language competence imposes some effects on the testees` performances, and the test results. They claimed that " the characteristics of the tasks used are always likely to affect assessment results to some degree, so that there is no assessment that yields only information about the ability want to assess" (p. 64).

Several studies have supported the role of test facets on the construct validity of the tests. Khodadady (2009) also performed a study on 38 senior students of English language teaching at Kurdistan University. The participants were selected both male and female, and their age ranged between 22-35 years. The researcher applied two tests in the study: 1) a traditional multiple-choice item test (MCIT), 2) a schema-based cloze-test. The first test content was selected from the passages which the students had covered during their course of study, and it consisted of sixty items. The deficiency of MCIT in the second test, according to Khodadady, was removed. By schema, the researcher, referred to meaning and semantic. The researcher designed a forty-item cloze with four schemata. He selected two passages from one of the students' books randomly. The designed items included four syntactic schemata: six adjectives, fourteen nouns, one adverb, and fifteen verbs. The two tests were administered in two sessions. At the end of the study, the researcher concluded that the reliability and internal validity of cloze-test was less than multiple-choice test, as the students were not familiar with the cloze passage and items opportunities. The students had to recall the answers of multiple-choice items to fill in the missing parts, but, in the cloze-test, the participants were able to respond the items as the test designer intended. In fact, the purpose was to find out how well the students comprehend the context to which the test designer intended. And, the results showed that answering multiple-choice test requires remembrance of what student have learnt from the given passage, however, schema-based cloze multiple-choice requires reading and understanding the given passage so that the participant select the target schema which was deleted from the text.

Rouhani (2008) conducted a study on 144 EFL students in two universities – Azad University of Khurasgan and Esfahan University- to investigate the validity and discrimination power of C-test on different language proficiency levels. He administrated two tests for two groups of one hundred and forty four EFL students. The first group consisted of 101 students whom received both a C-test and the Michigan test of English Language Proficiency from the same content, and the second group which was consisting of 43 students only received the C-test. The researcher concluded that C-test can be considered as an acceptable and suitable test in the case of construct validity. But, C-test did not appear to distinguish successfully between the participants of the two levels of lower and upper intermediate levels. It, also, failed to categorize the subjects according to their appropriate level of proficiency. However, the researcher found a high reliability in relation to C-test.

A. Statement of the Problem and Purpose of the Study

There were several studies in the case of effects of test facets on the construct validity of the tests. But most of these studies allocated a small part of their researches to examine the effects of test facets on construct validity comprehensively. Most of the pertinent studies lacked the completeness and the wide range of consideration of the effects of test methods on the construct validity of the tests. Another compelling issue is that most teachers prefer discrete-point items in testing because of ease of preparation, and ease of scoring; also most students prefer this type of testing because they are easy to be answered. So, designing and administering discreet-point tests, such as Multiple-choice and True/False will cause the students not to be able to answer the integrative-point tests such as Cloze-test and C-test skillfully from the same content. Although students have the similar competence from a passage, but if the content would be presented to students in different test facets, they may have difficulties in answering different forms of test items.

Taylor (2009) believes that "the familiarity or lack of familiarity with the content and the format of a test is acknowledged to impact on a test taker's performance, potentially leading to an over- or underestimation of the ability of interests" (p. 30). Moreover, integrative-point tests have minute share in the Iranian context. This form of evaluation should be thoroughly used in language testing in order to achieve precise information about students' competence. The main goal of the present study is to investigate the effects of test facets - Multiple-choice, True/False, Cloze-test and C-test- on the testees' performance. Besides, this study seeks to investigate the effects of four test forms on the construct validity of the tests, and to consider the testees' performance in relation to the way of test administration.

B. Research Question

The present study seeks to answer the following question:

1) Can test facets have impacts on Iranian EFL test takers' performances in terms of their proficiency level?

C. Research Hypothesis

The aim of the present study is to reveal the effects of test facets-integrative and discrete-point items- on the construct validity of the tests in Iranian EFL students. The results will, therefore, shed lights on the following hypothesis:

1) Test facets have impacts on Iranian EFL test takers' performances in terms of their proficiency level.

II. METHODOLOGY

A. Participants

The subjects of the study were students of three branches of Islamic Azad University in Khouzestan Province, -Dezful, Ahvaz, and Andimeshk. Totally, there were 160 students of 4 classes; one class was third year students at B.A. program of English Translation, another class was in the second year (last term) of A.A. program, and two classes in the first year of A.A to B.A. programs of English Language Teaching courses in the academic years of 2011-2012. The participants' age, both male and female, ranged from 20 to 34. Due to the unequally distributed pattern of students in terms of gender across the universities in the sixth region of the Islamic Azad University, the subjects were 114 female and 46 male English language learners.

To select a homogeneous sample, the researchers administered a proficiency test (Fowler & Coe, 1976) to 160 English language students studying at three branches of the university selected based on their availability. After taking the proficiency test, 160 students were categorized into three groups in terms of their proficiency level –upper-intermediate, intermediate, and low. The number of students in upper-intermediate, intermediate, and low levels were: 33, 100, and 27.

B. Instruments

Two main instruments used in the study were: 1) Nelson Proficiency Test (Fowler & Coe, 1976), 2) three multifaceted tests to test the three groups of proficiency levels. The researchers used the Nelson proficiency test to decide on homogenizing the selected sample and to specify the proficiency levels of the target sample in three levels. This test included 50 multiple-choice items, and each item was valued 1 point in scoring. 60 minute time span was allocated to test-takers to respond the test under controlled conditions so that participants make their real performance according to their English proficiency level. Those students whose scores fell within the range of +1 SD above and -1 SD below the mean were considered 'the intermediate level'. The scores which ranged below and above intermediate level were regarded as low and upper-intermediate proficiency level respectively. The numbers of participants in the low, intermediate, and upper-intermediate levels of proficiency were: 27, 100, and 33 respectively. The test included different items in English language so that they could cover all the English language elements including structural, lexical, and communicative aspects to show an appropriate image of participants' competence of English language. Also, to trust the reliability and appropriateness of this test to test the proficiency levels of selected samples, the researcher administered the test to 25 students of the selected sample as a pilot study beforehand. After taking the test, the researcher estimated the reliability of the gained data based on KR-21 formula to be 0.85.

After determining the English language proficiency levels of the participants –upper-intermediate, intermediate, and low, the researcher designed a multi-faceted test for each group of proficiency level. The questions were designed from a text extracted from "Exploring New Reading Strategies", volumes 1, 2, and 3 (Birjandi & Mosallanejad, 2010). Low level students were given the designed items from volume 1, intermediate level students took the designed items from volume 2 and finally the designed items from volume 3 were given to upper-intermediate students. So, each student received his/her special multi-faceted test based on the determined proficiency level – three multi-faceted tests for the three determined language proficiency levels.

C. Procedure

At the beginning of the study, Nelson Proficiency Test (Fowler & Coe, 1976) was administered to twenty five students from the selected sample to measure the reliability of the provided test items. The reliability of the test was 0.85. Afterwards, the sample took the test to check the difficulty level of the designed multi-faceted tests for each group of proficiency level. The provided time of taking the test was sixty minutes in which the testees were asked to answer 50 multiple-choice items of Nelson proficiency test. Having collecting the exam sheet of the proficiency levels were determined. Those testees whose scores were through the range of +1 SD above and -1 SD below the mean were considered the intermediate level. The scores which ranged below and above intermediate level were considered as low and upper-intermediate proficiency level, respectively. So, the number of participants in upper-intermediate, intermediate, and low level were 33, 100, and 27 respectively.

Then, the researchers administered the multi-faceted tests which were designed by the researchers themselves. First, the pilot study was conducted by the researchers to check the difficulty level of the designed multi-faceted tests for each group of proficiency level. So, the researcher selected 10 students- 5 male and 5 female. After collecting the test sheets, the difficulty level of the tests were calculated for each proficiency level. The goal of calculating such analysis was to see if the designed tests had appropriate difficulty level or not. According to Farhady, Ja'farpur, and Birjandi (2004), "items with facility indexes beyond 0.63 are too easy, and items with facility indexes below 0.37 are too difficult" (p. 101). So, the item difficulty level of each pilot group calculated by 1 minus item facility. The results showed that the designed multi-faceted tests for all three proficiency levels had acceptable difficulty, and there was no need to change or redesign the tests in order to be administered among testees. The item difficulty for low, intermediate and upper-intermediate levels were: 0.44, 0.41, and 0.40 respectively.

In the next step, the researchers implemented the multi-faceted tests for all the three proficiency levels. The testees of each proficiency level received the provided test of each level. Each testee was provided with a paper sheet which included a reading comprehension passage extracted from a text chosen from "Exploring New Reading Strategies", volumes 1, 2, and 3 (Birjandi & Mosallanejad 2010). The test items were designed from the volume 1 for low level, 2 for intermediate, and 3 for upper-intermediate level of proficiency. The test items included 5 multiple-choice items with three or four options, 5 true/false items, 5 items in matching cloze-test, and 5 items in c-test items facet. Participants had 30 minute time to respond to the test items. Also, attempts had been made that the administration procedure meets the requirements of a standard examination situation.

III. STATISTICAL ANALYSIS

Table 1 displays the mean, standard deviation and variance of participants` scores in Nelson proficiency test before the study and grouping. Table 1 shows the descriptive statistics of the Nelson proficiency test of 25 participants as a pilot study. Also it shows the statistics of this test for the whole samples under study-one hundred and sixty-.

TABLE1.					
DESCRIPTIVE STATISTICS OF PROFICIENCY TEST					
N Mean Std. Deviation Variance					
Sample	25	29.52	8.76	76.76	
Population	160	30.98	9.33	87.23	

Upper-intermediate level testees: Test form has an impact on the test takers' performance at upper-intermediate level. In other words, there are significant differences in the performances of language learners of upper-intermediate level as test-takers in different language tests forms. That is at Upper-intermediate level courses different language test forms lead to different performances on the learners' part. According to the obtained data presented in Table4.3, the mean of C-Test (4.09) is higher than the means of other test forms which are as follows/F (3.91), Cloze-Test (3) and MC (2.88).The descriptive statistics of the learners' performance of upper-intermediate level are presented in table2.

			I ABLE	. 2.	
DESCRIPTIVE	STATISTICS FC	or 33 stud	ENTS IN THE U	PPER-INTERMEDIATE	LEVEL OF THE PROFICIENCY
	Test	Ν	Mean	Std. Deviation	Std. Error
	M.C	33	2.88	1.083	.188
	True/False	33	3.91	.879	.230
	Cloze-test	33	3	1.323	.230
	C-test	33	4.09	.947	.165

For the comparison of the upper-intermediate level testees and the obtained data (achieved Means of the tests forms) to be drawn, an ANOVA analysis was used. The relevant result is 11.005 which at the level of P<0/001 can be considered significantly meaningful. As a result, it can be inferred that the observed F value is statistically significant.

		TABLE 3.			
	THE C	DNE-WAY AN	OVA		
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	37.909	3	12.636	11.005	0.001
Within Groups	146.970	128			
Total	184.879	131			

The obtained results confirm the difference among Upper-intermediate learners' performances in different language test forms. Finally, following the confirmation of existing the difference, Scheffe Test used in order to determine in which test forms the exact difference in performance exists. The results of Scheffe test are illustrated in table 4. The results of the Scheffe Test, as illustrated in table 4, show that there is a significant difference among the Upper-intermediate learners' obtained scores in test forms of MC, T/F and C-test items. Accordingly, it can be stated that Upper-intermediate learners did better at C-test items than those of the other test forms.

		SCHEFFE TEST: MULTIPLE COMPARISONS		
(I) Group	(J) Group	Mean Difference(I-J)	Std. Error	Sig.
M.C tests	True/False	-1.03*	.264	.002
	Cloze-Test	12	.264	.976
	C-Test	-1.21*	.264	.000
True/False	M.C Test Cloze-	1.03*	.264	.002
	Test	.91*	.264	.010
	C-Test	18	.264	.924
Cloze-Test	M.C Test	.12	.264	.976
	True/False	91*	.264	.010
	C-Test	-1.09*	.264	.001
C-test	M.C Test	1.21*	.264	.000
	True/False	.18	.264	.924
	Cloze-Test	1.09*	.264	.001

TABLE4. CHEFFE TEST: MULTIPLE COMPARISONS

*The mean difference is significant at the .05 level.

Intermediate level testees: Test form has an impact on the language learners' performances at intermediate level. In other words, there are significant differences in the performances of language learners of intermediate level in different language test forms. That is, at intermediate level courses different language test forms lead to different performances on the learners' part. Descriptive statistics of the intermediate level testees are shown in table5. According to the data illustrated in this table, it can be observed that T/F (4.22) has the highest Mean among other test forms which are C-Test (3.93), MC (3.06) and Cloze-Test (2.56), respectively.

TABLE5.					
DESCRIPT	IVE STATISTI	CS FOR 100 STU	JDENTS IN THE INTERMEDIATE LE	VEL OF THE PROFICIENCY	
Test	Ν	Mean	Std. Deviation	Std. Error	
M.C	100	3.06	1.229	.123	
True/False	100	4.22	.811	.081	
Cloze-test	100	2.56	1.653	.165	
C-test	100	3.93	.891	.089	

Table 6 shows that the observed F value equals 41.36 (P<0.001) which can be considered statistically significant. Based on the obtained data, it can be inferred that there is a significant difference among the intermediate-level learners' scores in different language test forms. To shed light on the issue, it can be stated that learners of intermediate level have achieved different scores on different test forms.

		TABLE 6.			
	THE C	DNE-WAY AN	OVA		
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	176.727	3	58.909	41.365	0.001
Within Groups	563.950	396	1.424		
Total	740.678	399			

As a result, following the confirmation of existing the difference in intermediate level testees' performance, Scheffe Test was used in order to determine where the exact difference lies. The results of Scheffe Test are illustrated in table 7. The results of Scheffe test, concerning the intermediate level testees' performance, show that there is a statistically significant difference between MC test forms scores and that of other test forms scores. In other words, the learners of intermediate level have scored differently in MC test form in contrast to other test forms (T/F, Cloze- Test). Also, there is a significant difference between the scores of T/F and Cloze-Test, and again between Cloze-Test and C-Test. The results show that the learners of intermediate level have performed better on T/F test forms except for C-test.

		TABLE7.				
SCHEFFE TEST: MULTIPLE COMPARISONS						
(I) Group	(J) Group	Mean Difference(I-J)	Std. Error	Sig.		
M.C tests	True/False	-1.16*	.169	.000		
	Cloze-Test	.50*	.169	.034		
	C-Test	87*	.169	.000		
True/False	M.C Test Cloze-	1.16*	.169	.000		
	Test	1.66*	.169	.000		
	C-Test	.29	.169	.400		
Cloze-Test	M.C Test	50*	.169	.34		
	True/False	-1.66*	.169	.000		
	C-Test	-1.37*	.169	.000		
C-test	M.C Test	.87*	.169	.000		
	True/False	29	.169	.400		
	Cloze-Test	1.37*	.169	.000		

*The mean difference is significant at the .05 level.

Low level testees. Test form has an impact on the language learners' performance at low level in a language test. In other words, there is a significant difference among the low-level learners' performances in different test forms. That is, at low level courses, different language test forms lead to different performances on the learners' part. It can be inferred from the data illustrated in table 8 that the mean of the T/F (4.26) is higher than that of other test forms which are as follows: C-Test (4.22), MC (3.85) and Cloze-test: (3.22).

			I ADLE 0.			
	DESCRIPTIVE STATISTICS FOR 27 STUDENTS IN THE LOW LEVEL OF THE PROFICIENCY					
Test	Ν	Mean	Std. Deviation	Std. Error		
M.C	27	3.85	.949	.900		
True/False	27	4.26	.944	.892		
Cloze-test	27	3.22	1.155	1.333		
C-test	27	4.22	.974	.949		

TADLES

Following that, a statistical analysis of the obtained Means and the obtained results of low level testees' performance are represented in table 9. As shown in table 9, the observed F value equals 6.133 (P<0.001) which is considered statistically significant. As a result, it can be inferred from the preceding data that there is a statistically significant difference among the low-level learners' performances in terms of different language test forms.

		TABLE 9.			
	THE C	DNE-WAY AN	OVA		
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	18.741	3	6.247	6.133	0.001
Within Groups	105.926	104	1.019		
Total	124.667	107			

Moreover, Scheffe test was applied in order to determine where the exact difference lies between the means. The results of the Scheffe test are presented in table 10. The results of the Scheffe test concerning the low level testees' performance show that there is a more significant difference among the low-level learners' performances in Cloze-test and C-test than that of other test forms. That is, they had poor performance in integrative tests. However, low-level learners performance in T/F test form.

TABLE 10.

SCHEFFE TEST: MULTIPLE COMPARISONS				
(I) Group	(J) Group	Mean Difference(I-J)	Std. Error	Sig.
M.C tests	True/False	41	.275	.534
	Cloze-Test	.63	.275	.161
	C-Test	37	.275	.613
True/False	M.C Test Cloze-	.41	.275	.534
	Test	1.04*	.275	.004
	C-Test	.04	.275	.999
Cloze-Test	M.C Test	63	.275	.161
	True/False	-1.04*	.275	.004
	C-Test	-1.00*	.275	.006
C-test	M.C Test	.37	.275	.613
	True/False	04	.275	.999
	Cloze-Test	1.00*	.275	.006

*The mean difference is significant at the .05 level.

IV. DISCUSSION AND CONCLUSION

According to the obtained data from the present study the researcher concluded that different test formats had impacts on the learners' performance variously at the three levels of language learning-intermediate to advance, intermediate and low. As the tables and the findings obtained suggest the students of intermediate level and upper-intermediate level had the poorest performance in MC test format, while they had done much better in C-Test, T/F, and Cloze-Test. A logical explanation for such phenomenon might be: Due to their vast knowledge of grammar and vocabulary the students of intermediate and advance level may suffer from some sort of misgiving as they find it somehow confusing to choose the correct option in MC test format. Moreover, there are some inevitable exceptions, for instance, a word with different meanings in different contexts, different word category, and different grammatical points, in any language which might be more of a burden on the students' part to distinguish the most suitable option and that results in hesitancy. As the time passes, intermediate and advance level students' knowledge and competence will develop to the point that they will make use of mental strategies in order to process the data and to respond the provided items in each test. Accordingly, as tests will be presented by some integrative test items in as much as they don't have any clue of the possible options, they will read the context, analyze it, and with the processes happening in their mind simultaneously they can find the answer without hesitation. But as mentioned before, students of intermediate and

advance level in discrete- point item tests had low performance because of their hesitation. Indeed, their high level of knowledge ability will drive them to make mistakes as they have to choose the most correct option among the correct options.

The testees' performance with different levels of language abilities and different background knowledge had been affected to some extent that they had different performances in different test formats. For instance, the intermediate and low level testees due to their unfamiliarity with Cloze-Test had poor performance in comparison to other test formats such as MC, T/F, and C-Test. They also had poor performance in comparison to upper-intermediate level testees in Cloze-Test. As mentioned before, the intermediate and low level testees' best performance was in MC, T/F, and C-Test where they were presented with some clues in the context of the test formats such as options or the first half of the stem, thus as they read the context or the options, they recall and choose the correct response. As mentioned before the poorest performance of the intermediate and low level testees was in Cloze-Test, as they had to process some mental abilities and strategies simultaneously in order to answer Cloze-Test items. These mental abilities and strategies which are reading, understanding, activating background knowledge-vocabulary, grammar, knowledge of world- and the ability of contextualization can be considered more of a burden on the testees' part due to their language abilities and their level of language learning as they had to process these mental abilities and strategies simultaneously.

That the upper-intermediate level testees have high performance in integrative tests such as C-Test and Cloze-Test is due to their high mental abilities and strategies that they process simultaneously. But their poor performance in MC tests can be sought in the nature of the MC test format itself, as it presents the testees with two, three, four or more alternatives to choose the correct response, thus it does not take the testees to go through the burden of processing and engaging mental strategies in order to find the correct response, it is just of a little of recalling and sometimes guessing. Moreover, the upper-intermediate level testees' poor performance in MC test format is, on one hand, due to their lack of attention to the clues provided in the stems and/or the other hand due to their over-attention to the provided clues which lead them to make mistakes and/or errors.

The most important outcome of the present study worth considering as it suggest measuring learners' language abilities based on different test formats, resulting in better judgments and decision making concerning the learners' performances. As it should be mentioned that one or two different test formats cannot be assumed reliable measures of the learners' language knowledge, basically, the reliability and validity of a test will be enhanced when the learners' language knowledge be measured based on different test formats, thus a fine judgment of their language abilities can be made.

REFERENCES

- [1] Bachman, L., & Palmer, A. (2010). Language assessment in practice. Oxford: Oxford University Press.
- [2] Baghaei, P. (2009). Understanding the Rasch model. Mashhad: Mashhad Islamic Azad University Press.
- [3] Birjandi, P. & Mosallanejad, P. (2010). Exploring new reading strategies. Tehran: Sepahan Publication.
- [4] Brown, J. D. (1995). Language program evaluation: Decision, problems and solution. *Annual Review of Applied Linguistics*, 15, 227-248.
- [5] Brown, J. D. (2000). Questions and answers about language testing statistics: What is construct validity?. JALT Testing & Evaluation SIG Newsletter, 4(2): 8-12.
- [6] Clapham, C. (2000). Assessment and testing. Annual Review of Applied Linguistics, 20, 147-161.
- [7] Cohen, A. D. (2001). "Second language assessment". In M. Celce-Murcia (Ed). Teaching English as a Second or Foreign Language, 515-534. Boston: Heinle & Heinle.
- [8] Cronbach, L., & Meehle, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- [9] Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Journal of Medical Education*, 37: 830-837.
- [10] Farhady, H., Ja'farpur, A., & Birjandi, P. (2004). Testing language skills from theory to practice (11th Ed). Tehran: The Center for Studying and Compiling University Books in Humanities (SAMT).
- [11] Fowler, W.S. & Coe, N. (1976). Nelson proficiency tests. London: Butler & Tanner Ltd.
- [12] Hughes, A. (2003). Testing for language teachers (2nd Ed). Cambridge: Cambridge University Press.
- [13] Khodadady, E. (2009). Objective-based achievement testing in the context of schema theory. *Iranian Journal of Language Studies* (IJLS), 3(1): 1-30.
- [14] Mahdavi-Zafarghandi, A., & Jahandideh, Y. (2006). An assessment of dentistry students` Reading proficiency of EST texts. Archive of SID, 18(15).
- [15] Pishghadam, R., & Tabataba`ian, M. S. (2011). Emotional intelligence: Can it be a predictor of performance on different test format?. *International Journal of Linguistics*, 3(1), 1-21.
- [16] Rouhani, M. (2008). Another look at the c-test: A validation study with Iranian EFL learners. *Asian EFL Journal*, Volume 10, Issue 1, Article 8.
- [17] Saeedi, M., Tavakoli, M., Rahimi Kazerooni, Sh., & Parvaresh, V. (2011). Do c-test and cloze procedure measure what they purport to be measuring?: A case of criterion-related validity. *International Journal of Human and Social Sciences*, 6:2, 99-108.
- [18] Shahivand, Z., & Pazhakh, A. R. (2012). The effects of test facets on the construct validity of the tests in Iranian EFL students. *Higher Education of Social Science*, 2(1): 16-20.
- [19] Taylor, L. (2009). Developing assessment. Annual Review of Applied Linguistics, 29, 21-36.



Zahra Shahivand is an instructor at Islamic Azad university of Iran, Andimeshk Branch. She has published several papers on testing and linguistics. Her research interests include testing language, sociolinguistics, and linguistics and discourse analysis.



Abbas Pazireshh was born on July 1, 1985 in Dishmook town of Iran. He achieved his BA translation studies from Islamic Azad University of Kazeroon, Iran. Morever, he got his MA degree in TEFL from Islamic Azad University, Science Research Branch, and Ahvaz, Iran. He is currently teaching at the department of translation studies of Payamnoor University of Iran, Dehdasht Branch. He is interested in performing and conducting research on different subjects such as discourse analysis, genre analysis, testing language, linguistics and predominantly sociolingustics studies. He is also the reviewig and editing member of Asian EFL journal and the director and founder of Serrat English institution in Dishmook. He has published several papers mostly on sociolinguistics internationally.



Alireza Raeeszadeh Is an instructor at Islamic Azad University of Ahvaz, Iran. His research favorites include psychology, neuro linguistics and linguistics.