

Emotional Intelligence: Can It Be a Predictor of Performance on Different Test Formats?

Reza Pishghadam (Corresponding author)

Dept. of English Language and Literature, Ferdowsi University of Mashhad, Iran

E-mail: pishghadam@um.ac.ir

Maryam Sadat Tabataba'ian

Dept. of English Language and Literature, Ferdowsi University of Mashhad, Iran

E-mail: m.tabatabaeyan@gmail.com

Abstract

This paper aims to investigate the relationship between Emotional Intelligence (EI) and test format in the light of test fairness considerations. This study took this relationship into account to see if people with different levels of EI performed differently on different test formats. To this end, 90 English language learners in Iran were asked to complete Bar-On's EQ test (1997) and a reading test which included four test formats (multiple choice, cloze test, c-test, summary writing). The results of the correlational analysis and regression equations indicate that EI constructs are in association with different test formats. Finally, the results were discussed in the context of language assessment and test fairness and some suggestions were made.

Keywords: Emotional intelligence, Test fairness, Test format, Language testing

1. Introduction

Test takers seem to perform differently on different test formats. Some prefer multiple-choice questions while others believe that they perform better on open ended questions. Among psychological factors that affect individuals' performance on different test formats, Emotional Quotient (EQ) seems to account for difference in performance on different test formats.

Bachman (2000) calls for the need for the investigation of the effect of certain characteristics on test taking process. These individual characteristics might affect different individuals differently and therefore, are a threat to the validity of the test. As assessment systems have been developed in order for a fairer selection to be done (Stobart, 2008), it is worth attending to the effect of individual characteristics on different test formats so that a fairer assessment can be achieved.

To the researchers' knowledge no study has been done to date to explore the relationship between EQ and test format. Therefore, the purpose of this study is to investigate the relationship between EQ and different test formats and to see whether EQ and its subscales can predict performance on different test formats.

1.1 Theoretical Foundations

In the following paragraphs, the notions related to test fairness, test format and emotional intelligence will be reviewed. First, test fairness is discussed as the findings of this study will contribute to this field and will provide test makers with another source that affects test fairness and test validity. The discussion of test format and emotional intelligence follows this part.

1.1.1 Test Fairness

Achieving a fairer assessment has always been one of educators' wishes although fairness cannot be thoroughly guaranteed. Shohamy (2001) states that testing must be continuously examined and this examination of testing must be done with care because tests are used for making high-stake decisions. Language tests like other tests are used for making decisions; so the tests used must be useful to both the society and individuals who live in the society (Fulcher & Davidson, 2007).

As testing plays an influential role in educational and social decisions about individuals, validation must take into account test's content, method and how test takers perform. As Messick (1989) states, social consequences of a test must be considered while validity framework is discussed (cited in Bachman, 1990).

The notion of validity has changed a lot after Messick's expanded view of validity was applied. Test fairness matters have been highlighted after this new notion of validity was proposed (Kunnan, 1999). In 1985, the unitary concept of validity was proposed and construct validity was seen to be the most important validity. Messick (1989) not only considered construct validity to be important, but also took test consequences into account (cited in Chapelle, 1999). Messick's paper (1989) drew the attention of scholars to the

consequences of the tests. The expanded notion of construct validity which Messick (1989) proposed includes both evidential and consequential bases for test validation (Kunnan, 1999). This framework is also believed to be a valuable basis for attending to issues of both validity and impact (Bachman, 2000).

Bachman (1990) also claimed that the inferences regarding specific uses of a test are validated, not the test itself. The use and interpretation of test performance may not be equally valid for all abilities and in all contexts. He also listed the sources of bias - cultural background, background knowledge, cognitive characteristics and native language, ethnicity, age and gender - which must be avoided. Differential validity takes test fairness into account. Attempts must be done to ensure that none of the sources of bias affect the measurement (Weir, 2005). It is important for the test not to be affected by the sources of bias as tests are used for making high-stakes decisions (Shohamy, 2001).

The aim of differential validity is to ensure that no testee will suffer because of the sources of bias and individual characteristics (Weir, 2005). According to Stobart (2005) effort must be made so that the tests are as fair as possible for the groups who want to take them.

1.1.2 Test Format

Many variables affect language test performance. One of these factors is test format. Bachman (1990) proposes a framework for test methods and revises it in Bachman and Palmer (1996). Bachman (1990) believes that test performance is affected by test method. Baker (1989) also believes that test method effect is important because we do not know whether test performance is due to the test takers' knowledge or their ability to answer certain formats.

According to Bachman and Palmer's framework (1996), method facets – or test task characteristics - can be divided into 5 categories:

1. Characteristics of setting;
2. Characteristics of the test rubrics;
3. Characteristics of input;
4. Characteristics of the expected response; and
5. Relationship between input and response.

The expected response, which is the focus of this study, can be of three types: selected response, limited production response, or extended production response (Bachman & Palmer, 1996). Brown and Hudson (1998) in the classification they propose classify language assessment into three broad categories:

1. Selected-response assessment
2. Constructed-response assessment
3. Personal-response assessment

The first type of assessment usually includes true-false, matching, and multiple-choice items. This type of test provides the language material for the student and the student is only expected to choose the correct answer from the available options. On the other hand,

constructed-response assessment contains fill-in, short-answer, and performance assessment. Unlike selected-response type of assessment, students are expected to produce language while doing this type of test. Finally, personal assessment includes conference, portfolio, and self- or peer-assessment. While students perform on this type of test, they actually produce language. The answers the students provide can be completely different and in this type of assessment the students communicate what they really want to communicate.

Many test formats have been employed in language testing, for example: cloze test, multiple-choice, c-test, gap-filling, matching, etc (In'nami & Koizumi, 2009). So far research done regarding different test formats has taken into account the relationship between field independence and language test performance (Hansen & Stanfield, 1981 & 1983), cognitive variables and language proficiency test performance (Chapelle & Roberts, 1986), test response format and text organization in reading comprehension tests (Kobayashi, 2002 & 2004), vocabulary test format and age (Bowels & Salthouse, 2008) and multiple-choice and open-ended test formats in L1 and L2 reading and L2 listening.

1.1.3 Emotional Intelligence

For years (1900-1969) emotion and intelligence were considered separate areas. Emotional intelligence first started in the field of academic psychology. In 1990-1993, Mayer and Salovey wrote a series of articles on EI and formed a formal theory of EI. They argued that EI was an overlooked intelligence. The first empirical measurement of EI was also developed. In 1994-1997, Goleman made it popular by writing his book *Emotional Intelligence* (Mayer, 2001).

EI is defined as:

“an ability to recognize the meanings of emotions and their relationships and to reason and problem-solve on the basis of them. Emotional intelligence is involved in the capacity to perceive emotions, assimilate emotion-related feelings, understand the information of those emotions, and manage them” (Mayer, Caruso & Salovey, 2000, p.267).

According to this definition of EI, information about relationships is one part of emotions and it is a standard intelligence that enriches discussions of human capacities (Mayer, Salovey, Caruso & Sitarenios, 2001). EI denotes the ability to process emotion-related information well and employing this information to guide cognitive activities (Salovey, Mayer & Caruso, 2002).

Goleman (1995) believes that people have two minds: the rational and the emotional. The emotional mind is the original and so he believes it matters more than IQ. He believes that a high IQ does not guarantee a good management of life instead the ability to control and use emotions is important.

Bar-On (1997) also stated that there are five broad areas relevant to success and these categories contain different subcategories. These 5 categories include:

1. Intrapersonal skills, which include emotional self-awareness, assertiveness, self-regard, self-actualization and independence.

2. Interpersonal skills, which include interpersonal relationships, social responsibility and empathy.
3. Adaptability scales, which include problem solving, reality testing and flexibility.
4. Stress management scales, which include stress tolerance and impulse control.
5. General mood, which includes happiness and optimism (cited in Mayer, Salovey & Caruso, 2000).

Research done in connection with EQ so far has paid attention to EQ as a predictor of success in transition from high school to university (Parker, Summerfeldt, Hogan & Majeski, 2004), and has investigated the way the constructs in emotional intelligence are related to a number of issues such as work settings (Carmeli, 2003), performance in interviewing (Fox & Spector, 2000), cognitive tasks (Shutte, Schuetplez, & Malouff, 2001), and foreign language learning (Pishghadam, 2009). Recently, some studies have been done into the role of emotional intelligence in teaching (Fer, 2004; Mortiboys, 2005; Hawkey, 2006). These studies have demonstrated that emotional intelligence can enhance the quality of teaching and learning. However, the relationship between EQ and testing has not been taken into account.

1.2 Purpose of the Study

As tests play an important role in teaching and as test fairness is one of the considerations of test developers, this study aims at seeking the relationship between EQ and test format. In this study, the aim is to focus on test formats used more often in different tests.

If a relationship is found, the tests must be reformatted. Test developers and teachers must change tests and all formats must be included in all different types of tests. This study is set out to answer the following three questions:

- Q1: Is there a significant relationship between EQ and performance of advanced learners on different test formats?
- Q2: Is there a significant difference between the means of high and low EQ groups, regarding performance of advanced learners on different test formats?
- Q3: Do EQ scores predict performance of advanced learners on different test formats significantly?

2. Method

2.1 Participants

The total population participated in this study consisted of 150 English learners from College of Ferdowsi University of Mashhad, Iran. Having been homogenized, only 90 learners were found to be at the same level of reading proficiency. Upper-intermediate and advanced students were chosen as they were able to speak and write English with a good command of grammatical structures and adequate vocabulary. As they were chosen from one institute, it could be ensured that all participants were familiar with the four test formats which were the concern of this study. To obtain more accurate results, the participants' reading comprehension

proficiency was homogenized using the reading part of standard paper-based TOEFL test administered in 2003. Out of the 90 participants, 50 were female and 40 were male. They were aged from 18 to 47, majoring in different fields of study at university.

2.2 Instrumentation

Two instruments were used to collect the data: Bar-On Emotional Quotient Inventory (EQ-i), and a reading test.

For measuring emotional intelligence, the Persian adaptation of Bar-On's EQ-i (Bar-On, 1997) was used, the short form of which was lately published as a 90 item questionnaire. Samouee (2002) has examined the reliability and validity of the Persian version of the questionnaire. Cronbach's Alpha exhibited a high reliability equal to 0.957 and its validity was substantiated through factor analysis. This inventory has five scales: (a) Intrapersonal, (b) Adaptability, (c) General Mood, (d) Interpersonal, (e) Stress Management. This questionnaire is scored according to a Likert-type scale of five points.

As EQ might be related to different skills, reading was chosen so that only the test format affects the results of the research, not the skill being tested. Two reading texts were chosen from paper-based TOEFL test which was administered in 2004. The two readings had topics which seemed to be familiar to all students in order for the test results not to be affected by some participants' topic familiarity. The response formats tested were also related to the same text so that different content does not affect test performance (Chen, 2004). Two tests were designed which tested four test formats: multiple-choice questions, cloze test, c-test and summary writing. Multiple-choice questions were chosen as the representative of selected response format as these questions have a fixed answer. The answer to c-test and cloze test is neither too fixed, nor too open-ended. Summarizing was also chosen as the most open-ended type of question with extended productive response.

The multiple-choice questions of the TOEFL test were used. Each reading had 10 multiple-choice questions so there were 20 multiple-choice questions to be answered (Cronbach's alpha = 0.523). The cloze test was prepared from the first half of the same reading texts. Every seventh word in the text was deleted (Farhady, Ja'farpur & Birjandi, 1994). The participants were expected to answer 50 cloze test questions (Cronbach's alpha = 0.568). The c-test was prepared from the second half of the same reading texts. Half of every other word was deleted (Raatz & Klein-Braley, 2002). The participants were expected to answer 50 c-test questions (Cronbach's alpha = 0.709). The students were also asked to write a summary of the reading texts in their own language. Summary writing was chosen instead of open-ended questions since as Kobayashi (2002) also states, summary writing measures overall comprehension better than open-ended questions. Like Kobayashi's study (2002), the participants were asked to write the summary in L1 so that the raters were not affected by their writing proficiency in English. Writing the summary in L1 also helped eliminate undesirable factors from performance on summary writing. Test takers could also refer back to the text so that the recall factor is eliminated. The summaries were rated by two raters and the interrater reliability was calculated to be 0.87. The test was piloted with 6 participants whose proficiency level was the same as the target population before it was given to the participants to make sure

that it was suitable for the students' level of proficiency (See Appendix).

2.3 Procedure

The process of data collection started in August, 2009 and ended in October, 2009. One hundred and fifty language learners, who were homogenized by the institute as they were studying at certain levels in the institute, took part in the reading part of an actual paper-based TOEFL test (2003) to ensure the homogeneity of the participants regarding their reading proficiency. There were five readings on the test and each had 10 questions so on the whole, the participants answered 50 multiple-choice reading questions. The test took about 90 minutes. The normal curve was drawn for the obtained grades and the students whose grades were within one standard deviation below and above the mean were chosen (N=90 learners).

When the participants were chosen, the reading test was given to them. The reading tests were administered in two sessions as each took about 70 minutes. The first parts of the test given to students were the cloze test and the c-test. It was given to the students first so that their memory did not affect test results. After that, they were asked to take the multiple-choice test and write the summary of the text. The EQ test was also given to the participants so that their EQ was obtained. The EQ-i is a self report, and it takes about 30 minutes to answer it. They did the test during a class session.

After collecting the data, it was entered into and processed with SPSS 17 program. The results gained from the tests taken by the participants fell within the interval data so the Pearson Product moment formula was used to calculate the correlation between each test format and EQ scores. According to the result of the participants' performance on the EQ test, two groups (high (N=45) and low (N=45)) were formed and the t-test was run to see whether the difference between the means of the high and low EQ groups was significant. Multiple regression analysis was also used to see which subscales of EQ were better predictors of performance on different test formats.

3. Results

The first question of this study was whether there was a relationship between test format and EQ. The following table shows the result of the correlational analysis.

Table 1. Correlation between test format and EQ construct

	Intrapersonal	Interpersonal	Adaptability	Stress	Mood	Total EQ
M.C.	.165	-.116	.238*	.165	.165	.170
Cloze Test	.186	-.122	.140	.136	.039	.123
C-Test	.274*	.079	.228*	.127	.154	.238*
Summary	-.063	-.246*	-.011	.120	-.100	-.066

As Table 1 presents, performance on multiple choice questions does not correlate with EQ. However, there is a moderate correlation between EQ, and c-test ($r = .238$, $p \leq .05$).

Considering the five subscales of EQ, it can be seen that Interpersonal Skills correlate negatively with summary writing ($r = -.246, p \leq .05$). It is also shown that Intrapersonal Skills correlate moderately with performance on c-test ($r = .274, p \leq .05$). Adaptability scales also have a moderate correlation with performance on multiple choice ($r = .238, p \leq .05$) and c-test ($r = .228, p \leq .05$).

To answer the second question t-test was also run to see if there existed a significant difference between the means of the high and low EQ groups. The results of the t-test study are shown in the following tables.

Table 2. Comparisons of performance on different test formats based on performance on EQ test

Variables	High EQ Group (n = 45) Mean	Low EQ Group (n = 45) Mean	T
M. C.	31.2778	29.2778	1.363
Cloze Test	21.8222	20.4000	1.390
C-Test	26.2667	24.5778	1.331
Summary	37.1944	37.6889	-.315

As Table 2 shows the difference between the high and low EQ groups is not significant so there is no significant difference between those with a high EQ and a low EQ regarding their performance on the four test formats.

3.1 Test performance in terms of Emotional Intelligence

This part shows the summary of the results obtained by multiple regression analysis using emotional intelligence as the predictor of performance on the four test formats.

Table 3. Prediction of Multiple-Choice by EQ

Predictors	R	R ²	Adjusted R ²	F	P	B
Interpersonal Adaptability	.341	.116	.096	5.715	.005	-.244 .267

Table 3 shows that Interpersonal Skills and Adaptability Scales of the EQ account for about 12% of the total variance in performance on multiple-choice test format ($R^2 = .116, p < .05$). Having a high Adaptability and low Interpersonal Skills were the best predictors of performance on multiple-choice questions.

Table 4 indicates that Interpersonal Skills and Intrapersonal Skills account for about 8% of the variance in performance on cloze test format ($R^2 = .075, p < .05$). Having high Intrapersonal Skills and low Interpersonal Skills were the best predictors of performance on cloze test questions.

Table 4. Prediction of Cloze Test by EQ

Predictors	R	R ²	Adjusted R ²	F	P	B
Interpersonal	.274	.075	.054	3.530	.034	-.135
Intrapersonal						.082

Table 5. Prediction of C-Test by EQ

Predictors	R	R ²	Adjusted R ²	F	P	B
Total EQ	.238	.057	.046	5.289	.024	.037
Intrapersonal	.274	.075	.065	7.156	.009	.106

As shown in Table 5 scores on EQ can predict about 6% of the total variance in performance on c-test questions ($R^2 = .057$, $p < .05$). It is also shown that Intrapersonal Skills account for about 8% of the total variance in performance on c-test format ($R^2 = .075$, $p < .05$). Having high EQ and high Intrapersonal Skills were the best predictors of performance on c-test questions.

Table 6. Prediction of Summary Writing by EQ

Predictors	R	R ²	Adjusted R ²	F	P	B
Interpersonal	.246	.060	.050	5.655	.020	-.235

As Table 6 indicates scores on Interpersonal subscale of EQ can predict about 6% of the total variance in performance on summary writing ($R^2 = .060$, $p < .05$). Having low Interpersonal Skills was the best predictor of performance on summary writing.

4. Conclusion

The present study sought to investigate the relationship between EQ and test format, to see if a significant difference between the means of the high and low EQ groups existed and finally to see how much EQ and its subscales predict performance on different test formats.

In the present study, it was shown that performance on multiple-choice questions is not related to EQ. However, Adaptability Scales were the only subscale of EQ that correlated with performance on multiple-choice questions. No correlation existed between EQ and its subscales and performance on cloze test. C-test was the only test format that correlated with EQ. Intrapersonal Skills also correlated significantly with performance on c-test. Summary writing also correlated negatively with Interpersonal Scales of the EQ test. The results of the t-test also indicated that the difference in means of the high and low EQ groups was not significant in any test format.

The obtained results from regression analysis indicate that EQ cannot predict performance on any test format except c-test. Out of the five subscales of EQ, Interpersonal Skills predicts performance on multiple-choice questions, cloze test and summary writing negatively. Adaptability Scales can only predict performance on multiple-choice questions. Finally, Intrapersonal Skills can predict performance on c-test and cloze test.

The findings show that EQ and its subscales are related to most test formats. It was formerly found out that multiple-choice questions also correlated with field independence and ambiguity tolerance (Chapelle and Roberts, 1986; Chapelle, 1988). In addition to field independence and ambiguity tolerance, Adaptability Scales of EQ also correlated with performance on multiple-choice questions. Adaptive individuals seem to be more successful in adapting themselves to multiple-choice questions.

Interestingly, Interpersonal Skills are negative predictors of performance on multiple-choice questions, cloze test and summary writing. Individuals with high Interpersonal Skills are good at communicating and understanding others so it seems they find it difficult to summarize information. Individuals who know themselves better (have high Intrapersonal Skills) seem to perform well on cloze test and c-test.

Surprisingly, EQ is only related to c-test and a person who is good in several subscales of EQ will be less successful in doing reading tests with certain formats so it is recommended that c-test be included in all reading tests.

The results of the present study are not in accordance with those done before on foreign language learning (Pishghadam & Ghonsooli, 2008; Fahim & Pishghadam, 2007; Pishghadam, 2009; etc). The reason might be that other studies have mostly taken grades of a term of study into account and students are usually graded based on both their performance and behavior in the classroom and their performance on written tests, while this study only considers the tests administered not how the test takers performed in the class during a term.

Alderson (2000) correctly points to the fact that employing only one method for measuring the understanding of the text is not adequate and objective and subjective methods of evaluation must be utilized side by side. He also states that good reading tests are the ones that use different techniques for assessing reading comprehension skills (cited in Weir, 2005).

This study has several implications for the test developers, teachers and researchers. As the results show, c-test is the only reading test format that is affected by EQ. However, different subscales of EQ are related to certain reading test formats and not others. Therefore, it is believed that different test formats must be included in reading tests so that people who are better in one of these subscales are neither advantaged nor disadvantaged.

As fairer assessment is the aim of testing and researchers all over the world are trying to develop fairer and more accurate tests, test developers must be careful not to include only one test format in the reading tests they develop so that the tests assess performance of test takers more accurately and more fairly. So, as Bachman (2000) also mentioned, language testers must start evaluating and analyzing tests with care to ensure test validity and they must make sure that the developed tests are ethical. The findings of this study help test developers develop reading tests that are free from sources of bias and treat test takers more equally.

Training of new testing professionals is also affected seriously by the findings of the present study. Testers can be trained who pay attention to all ethical notions and try to recognize sources of bias and avoid them.

National and international tests that are usually designed for selecting people for specific purposes and whose results are important to the people who take them must also be modified and the developers of such tests must take these differences in performance on different test formats into account so that the tests are fairer and they can be better measurements of competence than they are now.

The teachers must also be aware of these differences among their students so that they do not base their decisions only on the tests administered as these tests rarely include all test formats. Teachers as test users must understand the qualities of the tests they use and their appropriateness in the context these tests are used. They must try to develop fairer tests which include all test formats.

The false belief that standardized tests guarantee test fairness must be abandoned (Stobart, 2008). As Brown (2004) also questions standardized tests, they do not guarantee fairness for all participants and some will be in the minority group and if a more comprehensive picture of language learners' knowledge is required, different types of performance should be elicited.

If these differences are paid attention to, the designed tests will be more valid as the consequential validity of the tests, which is an important measure of validity and has been ignored until recently, will increase. And as Bachman (2000) correctly states, now that we have the methodological, theoretical and technological resources, plans must be made to ensure validity in practice and high quality tests must be developed.

As Bachman and Palmer (1996) also claim, as the effect of test task cannot be completely avoided, we must know them and understand them so that our tests are of high quality. The interpretations must also be done with care. As Bachman and Palmer (1996) correctly point out, the only characteristic of the test that is directly under our control is test task so care must be taken that it does not affect test performance.

The current study shows that test format affects reading test performance moderately so assessment must be done with care to ensure that this factor, which is irrelevant to language ability of test takers, does not affect test scores and their interpretation and in Messick's terms, construct- irrelevant variance is avoided.

As all sources of bias are not identified yet, it will be best to use a mixture of different test types. As Brown and Hudson (1998) correctly claim, when different test types are used, a more complete picture of the students' ability will be gained. As Hamp-Lyons (1997) states, testers must accept responsibility for all the consequences they are aware of (cited in Hamp-Lyons, 2000).

Shohamy (2000) also states that research in language teaching must inform testers of the areas that might cause problems for the test takers while performing on a test so that the tests are revised and a fairer and more accurate assessment can be achieved. That learners are not homogeneous groups must be attended to by test developers and teachers. So, both language testers and researchers must search for the effects of tests and the factors that affect test performance.

Stobart and Gipps (1998) believe that context and different types of response must be included in tests if we want to be fair (cited in Stobart, 2008). Shohamy (2001) also believes that tests that exclude certain groups of people must be omitted. She states that critical language testing favors the use of multiple procedures for assessing one's knowledge. In her opinion, tests have been originally developed for democratic purposes so they mustn't be used in undemocratic acts.

In the end, it is recommended that a more thorough attention be given to the factors that affect performance on different test formats. The same study with more participants and more reliable tests can be done. Different genders might also perform on different tests differently; therefore, it is worth considering its effect.

References

- Alderson, J. C. (1991). *Dis-sporting life. Response to Alistair Pollit's paper*. In Alderson and North (eds.), 60-67.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17, 1-42.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Baker, D. (1989). *Language testing: a critical survey and practical guide*. Great Britain: Hodder & Stoughton Limited.
- Bar-On, R. (1997). *Bar-On emotional quotient inventory: User's manual*. Toronto: Multi-Health Systems.
- Bowles, R. P., & Salthouse, T. A. (2008). Vocabulary test format and differential relations to age. *Psychology and Aging*, 23.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. USA: Longman.
- Brown, J. D., & Hudson, Th. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32, 4, 653-675.
- Carmeli, A. (2003). The relationship between emotional intelligence and work attitudes, behavior and outcomes. *Journal of Managerial Psychology*, 18, 788-813.
- Chapelle, C. (1988). Field independence: A source of language test variance? *Language Testing*, 5, 1, 62-82.
- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*. 19, 254-272.
- Chapelle, C., & Roberts, C. (1986). Ambiguity tolerance and field dependence as predictors

of proficiency in English as a second language. *Language Learning*, 36, 27-45.

Chen, L. (2004). On text structure, language proficiency, and reading comprehension test format interactions: A reply to Kobayashi, 2002. *Language Testing*, 21, 2, 228-234.

Fahim, M., & Pishghadam, R. (2007). On the role of emotional, psychometric, and verbal intelligences in the academic achievement of university students majoring in English language. *Asian EFL Journal*, 9.

Farhady, H., Jafarpur, A., & Birjandi, P. (1994). *Testing language skills: From theory to practice*. Tehran: The Organization for Researching and Composing University Textbooks for Humanities, SAMT.

Fer, S. (2004). Qualitative evaluation of emotional intelligence in-service program for secondary school teachers. *The Qualitative Report*, 9, 562-588.

Fox, S., & Spector, P.E. (2000). *Relations of emotional intelligence, practical intelligence, general intelligence, and trait affectivity with interview outcomes: It is not all just "g"*. *Journal of Organizational Behavior*, 21, 203-220.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: an advanced resource book*. USA: Routledge.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam.

Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical Concerns. *Language Testing*, 14, 3, 295-303.

Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System*, 28, 579-591.

Hansen, J., & Stansfield, C. (1981). The relationship between field independent-dependent cognitive styles and foreign language achievement. *Language Learning*, 31, 349-367.

Hawkey, K. (2006). Emotional intelligence and mentoring in pre-service teacher education: A literature review. *Mentoring & Tutoring*, 14, 137-147.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: focus on multiple-choice and open-ended formats. *Language Testing*, 26.

Kobayashi, K. (2004). *The effects of test methods on reading test scores of Chinese students learning Japanese as a foreign language*. Unpublished master's thesis. Ochanomizu University, Japan.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: text organization and response format. *Language Testing*, 19.

Kunnan, A. J. (1999). Recent developments in language testing. *Annual Review of Applied Linguistics*, 19, 235-253.

- Mayer, J. D. (2001). A field guide to emotional intelligence. In J. Ciarrochi, J. P. Forgas, & J. D. Mayer (Eds.), *Emotional intelligence in everyday life*. Philadelphia: Psychology Press.
- Mayer, J. D., Caruso, D. R., & Salovey, P. (2000). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, 27, 4, 267-298.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2000). *Models of emotional intelligence*. Cambridge, UK: Cambridge University Press.
- Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion*, 1, 3, 232-242.
- Messick, S. (1989). Validity. In R. L. Linn (ed.), *Educational measurement*. New York: Macmillan.
- Mortiboys, A. (2005). *Teaching with emotional intelligence: A step-by-step guide for higher and further education professionals*. London: Routledge.
- Parker, J. D. A., Summerfeldt, L. J., Hogan, M. J., & Majeski, S. A. (2004). Emotional intelligence and academic success: Examining the transition from high school to university. *Personality and Individual Differences*, 36, 163-172.
- Pishghadam, R. (2009). A Quantitative analysis of the relationship between emotional intelligence and foreign language learning. *Electronic Journal of Foreign Language Teaching*, 6, 31-41.
- Pishghadam, R., & Ghonsooli B. (2008). On the role of emotional intelligence in second language learning and academic achievement. *Studies in Foreign Languages*, 43, 41-56.
- Raatz, U., & Klein-Braley, Ch. (2002). Introduction to language testing and to c-tests. In J. A. Coleman, Grotjahn, Rudiger & U. Raatz (Eds.), *University language testing and the c-test*.
- Salovey, P., Mayer, J. D., & Caruso, D. R. (2002). The positive psychology of emotional intelligence. In C. S. Synder & Sh. J. Lopez (Eds.), *Handbook of positive psychology*. New York: Oxford University Press.
- Samou'i, R. (2002). *Emotional quotient inventory*. Esfahan: Ravan Tajhize Sina.
- Shohamy, E. (2000). The relationship between language testing and second language acquisition, revisited. *System*, 28, 541-553.
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18, 4, 373-391.
- Shuttes, N. S., Schuetplez, E., & Malouff, J. M. (2001). Emotional intelligence and task performance. *Imagination, Cognition, and Personality*, 20, 347-354.
- Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education*, 12, 3, 275-287.
- Stobart, G. (2008). *Testing times: the uses and abuses of assessment*. USA: Routledge.

Weir, C. J. (2005). *Language testing and validation: an evidence-based approach*. England: Palgrave.

Appendix

Appendix 1. Reading Tests

Cloze Test:

Read the text and complete it with the appropriate word.

After 1785, the production of children's books in the United States increased but remained largely reprints of British books, often those published by John Newbury, the first publisher to produce books aimed primarily at diverting a child audience. Ultimately, however, it was not the 1. _____, commercial-minded Newbury, but Anglo-Irish author Maria Edgeworth 2. _____ had the strongest influence on this 3. _____ of American children's literature. The eighteenth 4. _____ had seen a gradual shift away 5. _____ the spiritual intensity of earlier American 6. _____ writings for children, toward a more 7. _____ moralism. Newbury notwithstanding, Americans still looked 8. _____ children's books as vehicles for instruction, 9. _____ amusement, though they would accept a 10. _____ amount of fictional entertainment for the 11. _____ of more successful instruction. As the 12. _____ book market expanded, then, what both 13. _____ and publishers wanted was the kind 14. _____ fiction Maria Edgeworth wrote: stories interesting 15. _____ to attract children and morally instructive 16. _____ to allay adult distrust of fiction.

American 17. _____ against imported books for children set 18. _____ after the War of 1812 with the British. 19. _____ wave of nationalism permeated everything and 20. _____ self-conscious new nation found foreign writings (21. _____ those from the British monarchy) unsuitable 22. _____ the children of a democratic republic, 23. _____ state of self-governing, equal citizens. Publishers 24. _____ children's books began to encourage American 25. _____ to write for American children.

C-test:

Read the text and provide the second half of each word.

When they responded, the pattern established by Maria Edgeworth was at hand, attractive to most of them for both its rationalism and its high moral tone. Early 1.i ___ the 1820's, 2.sto ___ of 3.wil ___ children 4.tea ___ to 5.ob ___, of 6.care ___ children 7.lear ___ to 8.ta ___ care, 9.o ___ selfish 10.chil ___ learning 11.t ___ "live 12.f ___ others", 13.sta ___ to 14.fl ___ from American 15.pres ___, successfully 16.achi ___ Edgeworth's 17.to ___, though 18.rar ___ her 19.liv ___ style. 20.Imit ___ as 21.th ___ were, 22.t ___ early American 23.sto ___ were 24.qu ___ distinguishable 25.fr ___ their British counterparts. Few servants appeared in them, and if class distinctions had by no means disappeared, there was much democratic insistence on the worthiness of every level of birth at work. The characters of children in this fiction were serious, conscientious, self-reflective, and independent-testimony to the continuing influence of the earlier American moralistic tradition in children's books.

Multiple-choice Questions:

Read the reading and do the following exercises.

After 1785, the production of children's books in the United States increased but remained largely reprints of British books, often those published by John Newbury, the first publisher to produce books aimed primarily at diverting a child audience. Ultimately, however, it was not the cheerful, commercial-minded Newbury, but Anglo-Irish author Maria Edgeworth who had the strongest influence on this period of American children's literature. The eighteenth century had seen a gradual shift away from the spiritual intensity of earlier American religious writings for children, toward a more generalized moralism. Newbury notwithstanding, Americans still looked on children's books as vehicles for instruction, not amusement, though they would accept a moderate amount of fictional entertainment for the sake of more successful instruction. As the children's book market expanded, then, what both public and publishers wanted was the kind of fiction Maria Edgeworth wrote: stories interesting enough to attract children and morally instructive enough to allay adult distrust of fiction.

American reaction against imported books for children set in after the War of 1812 with the British. A wave of nationalism permeated everything and the self-conscious new nation found foreign writings (particularly those from the British monarchy) unsuitable for the children of a democratic republic, a state of self-governing, equal citizens. Publishers of children's books began to encourage American writers to write for American children. When they responded, the pattern established by Maria Edgeworth was at hand, attractive to most of them for both its rationalism and its high moral tone. Early in the 1820's, stories of willful children teaming to obey, of careless children learning to take care, of selfish children learning to "live for others", started to flow from American presses, successfully achieving Edgeworth's tone, though rarely her lively style. Imitative as they were, these early American stories were quite distinguishable from their British counterparts. Few servants appeared in them, and if class distinctions had by no means disappeared, there was much democratic insistence on the worthiness of every level of birth at work. The characters of children in this fiction were serious, conscientious, self-reflective, and independent-testimony to the continuing influence of the earlier American moralistic tradition in children's books.

Read the reading and check the correct answer.

1. What does the passage mainly discuss?
 - a. The career of Maria Edgeworth as an author of children's books
 - b. The development of children's literature in the United States
 - c. Successful publishers of children's books in Britain and North America
 - d. Basic differences between British and American literature for children
2. The publisher John Newbury is principally known for which of the following reasons?
 - a. He produced and sold books written by Maria Edgeworth.
 - b. He had more influence on American children's literature than any other publisher.

- c. He published books aimed at amusing children rather than instructing them.
 - d. He was commercially minded and cheerful.
3. The word “notwithstanding” in line 7 is closest in meaning to
 - a. in spite of
 - b. in addition to
 - c. as a result of
 - d. as a part of
4. The word “they” in line 9 refers to
 - a. children
 - b. Americans
 - c. books
 - d. vehicles
5. The word “allay” in line 12 is closest in meaning to
 - a. clarify
 - b. attack
 - c. reduce
 - d. confirm
6. It can be inferred from the passage that American children’s books sold before 1785 were almost always
 - a. written by Maria Edgeworth
 - b. attractive and interesting to children
 - c. written by American authors
 - d. intended only for religious and moral instruction
7. By the end of the eighteenth century, the publishers of children’s books in the United States were most concerned about which of the following?
 - a. Attracting children with entertaining stories that provided lessons of correct behavior
 - b. Publishing literature consisting of exciting stories that would appeal to both children and adults
 - c. Expanding markets for books in both Britain and the United States
 - d. Reprinting fictional books from earlier in the century
8. The word “permeated” in line 14 is closest in meaning to
 - a. opposed
 - b. improved
 - c. competed with
 - d. spread through
9. According to the passage, American children’s stories differed from their British

equivalents in that the characters in American stories were

- a. children who showed a change of behavior
- b. children who were well behaved
- c. rarely servants
- d. generally not from a variety of social classes

10. The word “testimony to” in line 26 is closest in meaning to

- a. inspiration for
- b. evidence of
- c. requirement for
- d. development of

Summary:

*Write a summary of the text you read in **Farsi**.*

Cloze Test:

Read the text and complete it with the appropriate word.

If food is allowed to stand for some time, it putrefies. When the putrefied material is examined 1. _____, it is found to be teeming 2. _____ bacteria. Where do these bacteria come 3. _____, since they are not seen in 4. _____ food? Even until the mid-nineteenth century, 5. _____ people believed that such microorganisms originated 6. _____ spontaneous generation, a hypothetical process by 7. _____ living organisms develop from nonliving matter.

8. _____ most powerful opponent of the theory 9. _____ spontaneous generation was the French chemist 10. _____ microbiologist Louis Pasteur (1822-1895). Pasteur showed that 11. _____ present in air closely resemble the 12. _____ seen in putrefying materials. He did 13. _____ by passing air through guncotton filters, 14. _____ fibers of which stop solid particles. 15. _____ the guncotton was dissolved in a 16. _____ of alcohol and ether, the particles 17. _____ it had trapped fell to the 18. _____ of the liquid and were examined 19. _____ a microscope slide. Pasteur found that 20. _____ ordinary air there exists a variety 21. _____ solid structures ranging in size from 0.01 mm 22. _____ more than 1.0 mm. Many of these bodies 23. _____ the reproductive structures of common molds, 24. _____ animals, and various other microbial cells. 25. _____ many as 20 to 30 of them were found in fifteen liters of ordinary air, and they could not be distinguished from the organisms found in much larger numbers in putrefying materials. Pasteur concluded that the organisms found in putrefying materials originated from the organized bodies present in the air. He postulated that these bodies are constantly being deposited on all objects.

C-test:

Read the text and provide the second half of each word.

Pasteur showed that if a nutrient solution was sealed in a glass flask and heated to boiling to destroy all the living organisms contaminating it, it never putrefied. The 1.propo___ of 2.spont___ generation 3.decl___ that 4.fr___ air 5.w___ necessary 6.f___ spontaneous 7.gener___ and 8.th___ the 9.a___ inside 10.t___ sealed 11.fl___ was 12.affe___ in 13.so___ way 14.b___ heating 15.s___ that 16.i___ would 17.n___ longer 18.sup___ spontaneous 19.gener___. Pasteur 20.const___ a 21.swan-n___ flask 22.i___ which 23.putre___ materials 24.co___ be 25.heat___ to boiling, but air could reenter. The bends in the neck prevented microorganisms from getting in the flask. Material sterilized in such a flask did not putrefy.

Multiple-Choice Questions:

Read the reading and do the following exercises.

If food is allowed to stand for some time, it putrefies. When the putrefied material is examined microscopically, it is found to be teeming with bacteria. Where do these bacteria come from, since they are not seen in fresh food? Even until the mid-nineteenth century, many people believed that such microorganisms originated by spontaneous generation, a hypothetical process by which living organisms develop from nonliving matter.

The most powerful opponent of the theory of spontaneous generation was the French chemist and microbiologist Louis Pasteur (1822-1895). Pasteur showed that structures present in air closely resemble the microorganisms seen in putrefying materials. He did this by passing air through guncotton filters, the fibers of which stop solid particles. After the guncotton was dissolved in a mixture of alcohol and ether, the particles that it had trapped fell to the bottom of the liquid and were examined on a microscope slide. Pasteur found that in ordinary air there exists a variety of solid structures ranging in size from 0.01 mm to more than 1.0 mm. Many of these bodies resembled the reproductive structures of common molds, single-celled animals, and various other microbial cells. As many as 20 to 30 of them were found in fifteen liters of ordinary air, and they could not be distinguished from the organisms found in much larger numbers in putrefying materials. Pasteur concluded that the organisms found in putrefying materials originated from the organized bodies present in the air. He postulated that these bodies are constantly being deposited on all objects.

Pasteur showed that if a nutrient solution was sealed in a glass flask and heated to boiling to destroy all the living organisms contaminating it, it never putrefied. The proponents of spontaneous generation declared that fresh air was necessary for spontaneous generation and that the air inside the sealed flask was affected in some way by heating so that it would no longer support spontaneous generation. Pasteur constructed a swan-necked flask in which putrefying materials could be heated to boiling, but air could reenter. The bends in the neck prevented microorganisms from getting in the flask. Material sterilized in such a flask did not putrefy.

Read the reading and check the correct answer.

1. What does the passage mainly discuss?
 - a. Pasteur's influence on the development of the microscope.

- b. The origin of the theory of spontaneous generation.
 - c. The effects of pasteurization on food.
 - d. Pasteur's argument against the theory of spontaneous generation.
2. The phrase "teeming with" in line 2 is closest in meaning to
 - a. full of
 - b. developing into
 - c. resistant to
 - d. hurt by
3. Which of the following questions did the theory of spontaneous generation attempt to answer?
 - a. What is the origin of the living organisms seen on some food?
 - b. How many types of organisms can be found on food?
 - c. What is the most effective way to prepare living organisms for microscopic examination?
 - d. How long can food stand before it putrefies?
4. The word "resemble" in line 8 is closest in meaning to
 - a. benefit from
 - b. appear similar to
 - c. join together with
 - d. grow from
5. The purpose of the "guncotton" mentioned in paragraph 2 was to
 - a. trap particles for analysis
 - b. slow the process of putrefaction
 - c. increase the airflow to the microscopic slide
 - d. aid the mixing of alcohol and ether
6. The author mentions "1.0 mm" in line 12 in describing the
 - a. thickness of a layer of organisms that was deposited on an object
 - b. diameter of the fibers that were in the guncotton filters
 - c. thickness of the microscopic slides that were used
 - d. size of the particles that were collected
7. The word "postulated" in line 17 is closest in meaning to
 - a. analyzed
 - b. doubted
 - c. persuaded
 - d. suggested
8. The word "it" in line 20 refers to

- a. a nutrient solution
 - b. a glass flask
 - c. boiling
 - d. spontaneous generation
9. According to paragraph 3, proponents of spontaneous generation believed that which of the following was important for the process to succeed?
- a. A sealed container
 - b. Fresh air
 - c. Heat
 - d. The presence of nutrients
10. It can be inferred from paragraph 3 that Pasteur employed a swan-necked flask to
- a. store sterilized liquids for use in future experiments
 - b. prevent heat from building up in a solution
 - c. disprove a criticism of his conclusions
 - d. estimate the number of organisms in a liter of air

Summary:

Write a summary of the text you read in Farsi.

Copyright Disclaimer

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).