An Analysis of the Relationship between Test Method, Personality Type, and Gender: A Case of the Iranian EFL Learners

Dr Parviz Birjandi *
Godrat H. Rostamlu *

Abstract

During the last few decades of the twentieth century, many research studies have been carried out in order to delve into the construct-irrelevant factors affecting language test performance of EFL/ESL learners. In line with this orientation and on the basis of some relevant assumptions made by scholars like Bachman (1990), Shohamy *et al.* (1993), Bachman *et al,* (1996), etc. , the study reported here was designed to examine the impact of test method, personality type and gender on the Iranian EFL learner's test performance and analyze the probable patterns of interaction between these variables in order to suggest appropriate test methods for different personality types and the two sexes and consequently maximize validity and minimize biasedness in language testing.

To fulfill the above-mentioned objective, a group of 125 Iranian EFL learners were asked to take three language measures exclusively developed for the present study (a multiple-choice test, an open-ended test and a cloze test), an oral interview and the Eysenck Personality Questionnaire.

* Allameh Tabatabaee University

Concerning test method effect, we could conclude that this variable per se does not significantly affect the Iranian EFL learners' test performance. Likewise, neuroticism was not found to be a significant factor in this respect. However, our data analysis results revealed that extroversion and sex positively affect the subject's test performance and psychoticism negatively do so.

With regard to the relationship between the above-mentioned variables, it was found that test method is significantly related neither to personality type nor to sex. We also could not find any interaction effect between personality type and sex in relation to different test methods.

I. Introduction

1. Background

It is generally believed that one of the characteristics of different disciplines including language teaching/learning in the modern age is attaching great importance to testing and measurement. This emphasis is reflective of a wide range of applications that testing and measurement have in different branches of science in general and in language teaching/learning in particular. Every day millions of educational and vocational decisions or inferences are made on the basis of a wide array of tests. Now, the important question to be considered is 'How accurate are the

• Imam Hosein University

inferences, descriptions, classifications, and decisions made about individuals on the basis of their test performance?'

A closer examination of testing literature reveals that current testing lacks precision and in a large number of cases leads to unfairness in the allocation of opportunities. Test results affected by numerous non-construct and external factors, in a lot of cases, mislead us as indicators of testees' abilities.

During the last two decades, many language testing scholars and researchers have thought over this situation and consequently a number of important and new lines of research have been set out in order to diagnose and remedy the problems negatively affecting the use of language tests. One important line of such research concerns ethics and fairness in language testing. Researchers attracted towards this issue (Elder, 1997; Bachman, 1990, 1991; Howe, 1995; Shohamy, 1997; Alderson and Banerjee, 2001; etc.) have endeavored to identify and measure sources of unfairness and therefore make the decisions made on the basis of language tests ethical. They have mostly focused on such variables as race, ethnicity, sex, scoring of responses, language background and educational background.

The second important line of recent research in the field of language testing is related to some new concerns in test validity theory. Having challenged the traditional 'trinitarian doctrine' of test validity (content validity, criterion-related validity and construct validity), Messick (1988a, 1988b)

presented a unitary concept of it. Messick (1996) mentions two standards for this unified model of validity: authenticity and directness. The concern of the former is that nothing important be left out of assessment of the focal construct and what the latter concerns is that "nothing irrelevant be added that interferes with or contaminates construct assessment." (p. 244)

On the basis of the assumptions made by Shohamy (1984), Backman (1990), Backman *et al.* (1996), etc., it was postulated that test method, personality type and sex may function as three non-target constructs and hence systematic sources of assessment errors contributing to test unfairness and invalidity (in Messichian sense) of language tests. It was also postulated that understanding the nature of these variables and identifying their extent of influence on test performance may help testers take a giant step towards developing fairer language tests and fulfilling Messick's standards of unified test validity.

In view of these postulations, the present study was conducted to investigate the following research questions:

1) Does test method significantly affect the Iranian EFL learners' test performance?
2) Does personality type (extroversion, psychoticism and neuroticism) significantly affect the Iranian EFL learners' test performance?
3) Is there a significant relationship between test method and the Iranian EFL learners' personality type?

4) Does sex significantly affect the Iranian EFL learners' test performance?

5) Is there a significant relationship between test method and the Iranian EFL learners' sex?

6) Is there a significant relationship between the Iranian EFL learners' sex, personality type and test performance?

2. Test Method Effect

Test method refers to the type of discourse included in the test or test questions. Bachman (1990) speaks about general test methods such as "performance test method" of which oral interview and essay are two examples. He also talks about specific test methods such as multiple-choice, completion and cloze. In this study the intended category of test method is the second one and the methods employed are multiple-choice, open-ended, cloze and oral interview.

Test method effect is a kind of bias resulted from various facets of testing method on the score received by a test taker in a specific language trait. This kind of bias emerges if, for example, an examinee's performance in a listening comprehension test is dependent to any extent on how such a trait is being measured, e.g. by open-ended questions, multiple-choice items, etc. and thus the score he obtains is not really a reflection of his knowledge on listening comprehension trait.

Influence of test method on language test performance does not seem to be a strange mechanism. It's a long time

that researchers have found out different individuals utilize different strategies in processing and learning language materials. Now seeing that discourse type of test questions arouse different reactions on the part of test takers and influence their performance should not surprise us.

The effect of test method on learner performance has been underlined by many researchers. Alderson (1991) says "research has clearly shown that there is such a thing as method effect in testing... and it is important that we minimize method bias in our measurement". (p. 11) Shohamy (1997) contends "in studies on test method, it was found that the test itself affects the scores that test takers obtain on tests".(p.341)

3. Personality Type

In this study, definition of personality type originates from The Eysenck Personality Questionnaire (EPQ) which is constructed on the basis of Hans Eysenck's trait theory of personality.

The trait personality theory is founded according to the conception that the same stimulus or social event often raises different responses on the part of different individuals. This theory also assumes that every individual's personality consists of a number of personality characteristics or traits which categorize him/her according to the degree to which he/she manifests them.

Hans Eysenck, whose trait personality theory has served as the basis for the present study, believes that

personality is that which permits a prediction of human behavior in a given situation. In Eysenck's theory, there are three dimensions or supertraits placed at the highest level of his hierarchical model of personality. As Burger (1990) explains, in this hierarchical scheme the basic structure is the *specific response* level, which includes specific behaviors. For example, if a man is seen spending the afternoon talking and laughing with friends, he is displaying a specific response. If the same individual spends many afternoons having a good time with friends, we have evidence for the second level in Eysenck's scheme, *a habitual response.* It is likely that the man doesn't limit himself to socializing just in the afternoons and just with these friends. He may devote a large part of his weekends and quite a few evenings to his social life participating in social gatherings, discussion groups, parties, and so on. Under such conditions, according to Eysenck, this man exhibits the *trait* of sociability. Eysenck believes traits such as sociability are part of a still larger personality dimension. It means that a sociable person also tends to be impulsive, active, lively, and excitable. All these traits combine to form the *supertrait* called **extroversion**.

Early in his research, Eysenck found two supertraits: extroversion/introversion(E) and neuroticism(N). In 1952, Eysenck hypothesized that a third supertrait or major dimension of personality exists independent of extroversion/introversion and neuroticism and he called it psychtocism(P).

Eysenck and Eysenck (1991) say the typical extrovert "is sociable, likes parties, has many friends, needs to have people to talk to, and does not like reading or studying by himself". They also say that a high N scorer is an anxious, worrying individual who is moody and frequently depressed. He is likely to sleep badly and to suffer from various psychological disorders. He is overly emotional, and too quickly reacts to all sorts of stimuli. In Eysenck's words, a high P scorer is a solitary type, not caring for people; he is often troublesome, not fitting in anywhere. He may be cruel and inhumane, lacking in feeling and empathy, and altogether insensitive. He has a liking for odd and unusual things and a disregard for danger.

According to Eysenck and Eysenck (1991), these three psychiatric terms should not be taken to imply that the scales are not useful for the measurement of personality traits in normal persons. They further say these terms refer to underlying dispositional personality traits which are present in varying degrees in all persons. Thus, only a small proportion of people with high P, for example, are likely to develop a psychosis (an abnormal condition) in the course of their lives. Therefore, in order to avoid any misunderstanding and for many practical purposes, Eysenck and Eysenck (1991) recommend that we omit psychiatric terms like 'neuroticism' and 'psychoticism' and replace them by terms like 'emotionality' and 'tough-mindedness' respectively.

Among the three above-mentioned personality types, extroversion/introversion is the most frequently investigated one. Concerning the effect of this bipolar supertrait on language test performance and its relationship with different language constructs, three patterns of research findings are discernable in the literature. Researchers like Swain and Burnaby (1976), Sutter (1976), Naiman *et al.* (1978), etc. have not found extroversion/introversion an effective factor in EFL/ESL learning. Busch (1982), Naylor (1982), cited in Abbssy (1995), Astika *et al.* (1996) etc. have pointed to a negative relationship between extroversion and EFL/ESL learning. And such researchers as Brown (1973), Chastain (1975), MacIntyre and Charos (1996), etc. have come up with positive relationship between extroversion EFL/ESL learning.

4. Gender

With respect to the role of sex in EFL/ESL learning, there is a commonly-held belief to the effect that girls are in an advantageous position. However, a review of literature indicates that the reality of the case is not as straight-forward as this. The findings of the related empirical research are quite disparate and as Takala and Kaftandjieva (2000) put it "the research findings differ significantly from statements such as girls have greater verbal ability (Maccoby and Jacklin, 1974; Elwood, 1995; Cole, 1997) through 'there are no general differences in verbal ability (Hyde and Lynn, 1988) to women obtained lower means

than men on the verbal scale (Lynn and Mulhern, 1991; Lynn and Dai, 1993; Born and Lynn, 1994)". (pp. 323 and 324)

Whatever role sex plays in EFL/ESL learning, with respect to measuring language ability it is said to function as a source of bias. This phenomenon occurs when men and women with the same ability levels tend to obtain different test scores. The following factors are believed to cause this problem:

1)The test method employed may be favored by one gender.

2)The material or content of a test may be offensive to members of one gender.

3)The test content may be more familiar to men or women.

4)The rater or raters may be biased towards men or women.

5)Men and women may be unequally represented as actors in test items.

6)Men and women may have special reactions toward their interlocutors in a test

method like interview.

From among these factors, the role of the first one is intended to be examined in the present study.

II. Method

1. Subjects

The subjects taking part in this research project were a group of 125 learners of English as a foreign language at Jihade-Daneshgahi Institute of Tehran University which is a language institute offering a number of non-degree English language courses. The aforementioned number of subjects were studying English at upper-intermediate level and their textbook was *Headway Upper-intermediate* by John and Liz Soars. The participants were not randomly selected and the whole group came from a number of intact classes. The native language of more than ninety percent of the subjects was Persian. Their age range was 16-40 with the average age of 26. The female subjects outnumbered the male ones by a rough proportion of 2 to 1 (82 females and 43 males).

2.Instruments

The necessary data for the present study were collected by making use of a multiple-choice test, an open-ended test, a cloze test, an oral interview, and a personality inventory. The first three measures were exclusively developed and standardized for the present study. The first two 50-item tests were constructed on the basis of a semester-long integrated course of 4 hours a week, and they were intended to evaluate the learners' achievement on listening comprehension, reading comprehension, grammar, and vocabulary. The cloze test was fixed-ratio with every eighth word deleted. The function of this test was to evaluate general language achievement of the subjects studying

English as a foreign language at uper-intermediate phase. The instrument employed for measuring the oral language proficiency of the subjects was the oral testing procedure developed by the Foreign Language Institute (FSI). The main objective of this measurement tool is to determine the level of speaking proficiency of candidates. And finally the Persian adult EPQ (1982) which is a personality inventory comprising extroversion/introversion, neuroticism or emotionality, psychoticism or tough-mindedness, and lie scales was used for assessing personality type of the subjects.

3. Procedures

To collect the necessary data for the present study, the five above-mentioned instruments were administered to the subjects during a language-school term of two and half months. The measures were administered during the subjects' regular class hours and under the supervision of their respective teachers.

III. Results and Discussion

To determine whether the six null hypotheses of the present study are rejected or confirmed, a number of relevant statistical analyses were applied on the collected raw data. Here, the results of these statistical procedures

will be presented and some light will be shed on their importance.

Hypothesis 1

Different test methods do not significantly affect the Iranian EFL learners' language test performance.

To test this null hypothesis, we submitted the relevant data to a one-way analysis of variance (ANOVA) the results of which showed that the methods employed were not significantly different for the subjects at $p<.05$ level of confidence. Therefore, the first null hypothesis was confirmed.

Table 1 One-way ANOVA of the mean scores on different test methods

	Sum of squares	df	Mean square	F	Sig.
Between groups	236.519	3	78.840	.356	.784
Within groups	53079.648	240	221.162		
Total	53316.167	243			

This finding implies that different test methods do not significantly affect the Iranian EFL learners' test performance. In light of the findings of some other research studies, (Shohamy, 1983; Shohamy *et al.*, 1986, Stansfield and Kenyon, 1981; Shohamy and Stansfield, 1991; etc.) which almost unanimously point to significant effect of test method on test performance, our finding on the first null hypothesis was somewhat unexpected although it provides additional support for the similar finding that Hesamy

(1994) had come up with. This unexpected finding seems to have been the function of the sample size which was relatively small in the case of the first hypothesis (only 61 students' had been use here). Had we employed a different sample – much larger in size and with much wider range of language proficiency, we might have come up with results much nearer to the already well-established findings related to the effect of test method on test performance.

Hypothesis 2

Personality type does not significantly affect the Iranian EFL learners' language test performance.

To test this hypothesis, a set of t-tests were run in order to compare the mean scores between the extroverts and the introverts, the neurotics and the non-neurotics, and the psychotics and the non-psychotics.

Table 2 Mean scores of different personality types on the four tests

Test	E	Mean	P	Mean	N	Mean
Open-ended	-E	49.97	-p	62.55	-N	56.99
	+E	66.68	+p	51.92	+N	53.46
Multiple-choice	-E	49.67	-p	62.55	-N	56.87
	+E	66.85	+p	51.70	+N	53.31
Cloze	-E	51.11	-p	62.84	-N	56.35
	+E	64.15	+p	51.30	+N	55.92
Interview	-E	56.61	-p	70.59	-N	67.20
	+E	74.80	+p	61.78	+N	64.73

Table 3 Results of t-tests on extroversion and test performance

	T-test for equality of means		
	T	df	Sig.(2-tailed)
Open-ended	-5.574	123	.000
Multiple-choice	-5.995	123	.000
Cloze	-4.675	123	.000
Interview	-6.712	59	.000

Table 4 Results of *t-tests* on psychoticism and test performance

	T-test for equality of means		
	T	df	Sig.(2-tailed)
Open-ended	3.360	123	.001
Multiple-choice	3.546	123	.001
Cloze	4.127	123	.000
Interview	3.177	59	.002

Table 5 Results of *t-tests* on neuroticism and test performance

	T-test for equality of means		
	T	df	Sig.(2-tailed)
Open-ended	.884	123	.378
Multiple-choice	.919	123	.360
Cloze	-.117	123	.907
Interview	.884	59	.378

A cursory look at tables 2, 3, 4 and 5 indicates that neuroticism does not have any significant effect on test performance, but the other two personality traits, i.e. extroversion and psychoticism, exert statistically significant influence on the Iranian EFL learners' test performance. It is also understood that the influence of extroversion is positive, but psychoticism negatively affects the subjects' test performance.

With respect to extroversion, our findings on this hypothesis confirm the general assumption held by the applied linguists to the effect that extroverts are in a more advantageous position in EFL/ESL learning. These findings are also in line with the most predominant research pattern discernible from the investigations in the field of applied linguistics (Brown, 1973; Chastain, 1975; Dewacle, 1993; Abbasy, 1995; etc.).

With regard to the findings related to psychoticism, we can say they were not unexpected in light of the suggestion made by Gibson (1981). According to this scholar, at trait level 'lack of empathy and solitude' are strongly associated with psychoticism. Some researchers like Guiora *et al.* (1967) have already concluded that 'empathy', which according to Larsen-Freeman and Long (1993) 'relates to an individuals' ability to put oneself in another's place', strongly contributes to language learning success in general and mastery of pronunciation accuracy in particular. Therefore, our results from hypothesis two denoting that psychotic language learners lacking in empathy are less successful than non-psychotic ones are compatible with the already-established research findings. With regard to 'solitude' as another trait involved in psychoticism, we can say in TEFL/TESL literature it is never associated with success, but rather it implies failure specially in acquiring oral language skills. Thus, our findings attesting to significantly better performance of non-psychotic subjects (with no degree of solitude) provide additional support for

the general assumptions held in the field of applied linguistics.

As for the effect of neuroticism on test performance, we should say the obtained result reveals a complex picture of the issue and we really expected statistically significant difference to the advantage of the non-neurotic subjects. Nevertheless, the only conceivable interpretation with regard to better performance of the non-neurotic subjects across the sample is that, according to Eysenck and Eysenck (1968), neurotic individuals mostly tend to be emotionally overresponsive and to have difficulties in returning to a normal state after emotional experiences, which seem to create serious difficulties for learning.

Hypothesis 3

There is no significant relationship between the Iranian EFL learners' personality types and their performance on different test methods.

For testing this hypothesis, three one-way ANOVAs were utilized the results of which are reported in the following three tables.

Table 6 One-way ANOVA of the extroverts' means scores on the four test methods

	Sum of squares	Df	Mean square	F	Sig.
Between groups	1899.710	3	633.237	2.262	.083
Within groups	45343.628	162	279.899		
Total	47243.337	165			

Table 7 One-way ANOVA of the psychotics' mean scores on the four test methods

	Sum of squares	Df	Mean square	F	Sig.
Between groups	2889.745	3	963.248	4.026	.008
Within groups	59806.851	250	239.227		
Total	62696.595	253			

Table 8 One-way ANOVA of the neurotics' mean scores on the four test methods

	Sum of squares	Df	Mean square	F	Sig.
Between groups	162.573	3	54.191	.239	.869
Within groups	20388.885	90	226.543		
Total	20551.457	93			

It is seen that only with respect to sychoticism the value of F is statistically significant which implies that there is a significant relationship between pschoticism and at least one of the test methods. This ANOVA result entailed looking for follow-up statistics in order to determine more precisely to which method or methods the independent variable (psychoticism) was related. For this purpose, we decided to run the Scheffe test the results of which are reported below.

Table 9 Multiple comparisons (Scheffe test results on the relationship between psychoticism and test methods)

(I) tests	(J) tests	Mean	Sig.

		difference (I-J)	
Open- ended	Multiple-choice	.2162	1.000
	Cloze	.6149	.996
	Interview	-9.8623*	.030
Multiple-choice	Open-ended	-.2162	1.000
	Cloze	.3986	.999
	Interview	-10.0785	.025
Cloze	Open-ended	-.6149	.996
	Multiple-choice	-.3986	.999
	Interview	-10.4772*	.018
Interview	Open-ended	9.8623*	.030
	Multiple-choice	10.0785*	.025
	Cloze	10.4772*	.018

* The mean difference is significant at the .05 level

The results of our post hoc comparisons (Table 9) revealed that the difference between the three pairs of means (interview vs. open-ended, interview vs. multiple-choice and interview vs. cloze) was significant for other than chance reasons with at least 95% certainty ($p<.05$). This finding leads us to conclude that our independent variable (psychoticism) significantly differs in how it is related to performance on the four test methods. It can be further said that such a difference and significant relationship exists between psychoticism and performance on the interview method.

We believe this finding is only attributable to the sample nature otherwise it cannot be justified in any conceivable way, because, in the first place, psychotic individuals with a high degree of solitude and lacking in empathy can never be apt to show an impressive performance on interview;

secondly, such a finding contradicts many well-established research findings including those of the present study attesting to the observation that extroversion greatly contributes to oral language skills as the traits making extroversion and psychoticism are almost diametrically opposed; and thirdly, to the researchers' knowledge, the finding is not grounded in any well-established theory in the field of applied linguistics.

Hypothesis 4

Sex does not significantly affect the Iranian EFL learners' language test performance.

For testing this hypothesis, t-test procedure was utilized. The findings of these analyses are reported in the following two tables.

Table 10 Descriptive statistics of the male and female subjects on the four tests

	Sex	N	Mean	Std. deviation	Std. error mean
Open-ended	Male	45	49.11	18.51	2.76
	Female	80	60.27	16.66	1.86
Multiple-choice	Male	45	49.69	17.50	2.61
	Female	80	59.75	16.65	1.86
Cloze	Male	45	50.50	15.02	2.24
	Female	80	59.11	16.30	1.82
Interview	Male	6	77.67	6.06	2.47
	Female	55	62.58	13.51	1.82

Table 11 Results of *t-tests* on sex and language test performance

	Levene's test for	T-test for equality of

1	equality of variances		means		
	F	Sig.	t	df	Sig. (2-tailed)
Open-ended	.243	.623	-3.454	123	.001
Multiple-choice	.526	.470	-3.183	123	.002
Cloze	.133	.716	-2.916	123	.004
Interview	3.210	.078	2.689	59	.009

It is seen that all four *t* values are high enough to let us assume that the male and female subjects had significantly different test performance and a cursory look at table 10 reveals that, except for the interview in which only 5 male subjects had participated, the difference is to the advantage of the females. Therefore, we can strongly reject the fourth null hypothesis and considering the limitations of the study claim that the Iranian females are more successful than males in learning English as a foreign language.

Although there are some research findings which contradict this result (Gross, 1983; Spurling and Illyin, 1985; Valizadeh, 1999; etc.), the majority of the researchers investigating the role of gender in EFL/ESL learning have found females superior to males (Farhady, 1982; Boyle, 1987; Eisenstein, 1989; Hyde and Lynn, 1988, cited in Takala and Kaftandjieve, 2000; etc.).

Hypothesis 5

There is no significant relationship between the Iranian EFL learners' sex and their performance on different test methods.

This hypothesis was also tested through one-way analysis of variance.

Table 12 One-way ANOVA of the females' means scores on the four test methods

	Sum of squares	df	Mean square	F	Sig.
Between groups	422.616	3	140.872	.549	.649
Within groups	7470.319	291	256.599		
Total	75092.936	294			

The results in this table indicate that the female subjects' mean scores on the four test methods are not significantly different at $p<.05$ level [F (3, 391)=.549, p=.649]. Therefore, it can be said that, as was the case with extroversion and neuroticism, use of one test method instead of another will not make any significant difference for the Iranian female learners of English as a foreign language.

Table 13 One-way ANOVA of the males' means scores on the four test methods

	Sum of squares	df	Mean square	F	Sig.
Between groups	4515.528	3	1505.176	5.332	.002
Within groups	38671.922	137	282.277		
Total	43187.450	140			

Although the F-observed here was high enough to let us think of significance of difference among the four mean scores, we thought the observed significance of difference was the function of the interview sample size which was much smaller and hence completely different from the other three in nature. To examine this hunch, we decided to put the interview mean score aside and compare the other three through a one-way ANOVA the results of which revealed that our hunch was right (Table 14). Therefore, as was the case with females, the male subjects' performance on different test methods does not show any significant difference and hence the methods under discussion equally suit the Iranian male learners of English as a foreign language.

Table 14 One-way ANOVA of the males' mean scores on three test methods

	Sum of squares	df	Mean square	F	Sig.
Between groups	43.811	2	21.906	.075	.928
Within groups	38488.589	132	291.580		
Total	**38532.400**	**134**			

Given the results of our two phases of analyses here, hypothesis five to the effect that there is no significant relationship between the Iranian EFL learners' sex and their performance on different test methods is supported. We can, therefore, claim that different test methods are equally suitable for the Iranian male and female learners of English as a foreign language.

Hypothesis five was proposed on the basis of the assumption that a test as a whole or a test method may be gender-biased and consequently lead to unfair estimation of EFL learners' latent ability. This being the case could lead to violation of Messick's validity criterion of 'directness' and hence invalidity of score inferences about target constructs. Therefore, we thought if a particular test method is found to be strongly related to females or males, this may help testers minimize bias in language assessment and consequently lead language testing discipline along a more constructive direction.

However, the statistical analyses on hypothesis five did not bear any evidence to substantiate our assumption. In other words, our findings on this hypothesis did not point to any relationship between sex and language test method. Although this result corresponds to the conclusion drawn by Hesamy (1994), it should not be taken as an undisputable fact. It needs to be further probed into in wider contexts and at larger scales.

Hypothesis 6

There is no significant relationship between the Iranian EFL learners' sex, personality type and test performance.

It needs to be emphasized that this hypothesis is not essentially concerned with examining the main effect of sex and personality type on test performance as this prime objective was taken care of in hypotheses two and four. What it intends to probe into is the interaction effect

between these two independent variables in relation to test performance. In order to fulfill this objective, we submitted the relevant raw data to two-way analysis of variance (factorial ANOVA). Since the subjects' test performance was available on four tests, we had to run four two-way ANOVAs. The following tables display the results of these analyses.

Table 15 Two-way ANOVA: sex by type in relation to the open-ended test

Source	Type III. sum of squares	df	Mean square	F	Sig.
Sex	3563.149	1	3563.149	13.609	.000
Type	7793.818	2	3896.909	14.883	.000
Sex by Type	366.032	2	183.016	.699	.499
Total	525844.000	147			

Table 16 Two-way ANOVA: sex by type in relation to the multiple-choice test

Source	Type III sum of squares	df	Mean square	F	Sig.
Sex	3176.229	1	3176.229	13.021	.000
Type	8211.342	2	4105.671	16.831	.000
Sex by type	340.473	2	170.237	.698	.499
Total	521068.000	147			

Table 17 Two-way ANOVA: sex by type in relation to the cloze test

Source	Type III sum of squares	df	Mean square	F	Sig.
Sex	2201.946	1	2201.946	10.010	.002
Type	5481.916	2	2740.810	12.460	.000
Sex by type	372.432	2	186.216	.847	.431

Total	**504689.750**	**147**

Table 18 Two-way ANOVA: sex by type in relation to the interview

Source	Type III sum of squares	df	Mean square	F	Sig.
Sex	**727.287**	**1**	**727.287**	**5.971**	**.017**
Type	**2061.726**	**2**	**1030.863**	**8.464**	**.001**
Sex by type	**267.411**	**1**	**267.411**	**2.196**	**.143**
Total	**320035.000**	**73**			

As expected, in all four ANOVA tables, the effect of sex factor exceeds the critical value and it is significant at $p<.05$ level of confidence. This finding is an additional support for the results on hypothesis four and, in practical terms, together with the information in the four above-presented tables of descriptive statistics, it implies that, except for the interview in which only five male subjects had participated which in turn makes any meaningful comparison impossible, the female EFL learners outperformed their male counterparts across the sample.

As for personality type, it is also seen that the main effect of this factor is significant which implies that subjects with different personality types have had differential performance across the sample. This result is also supported by the findings on hypothesis two.

However, a quick look at the four ANOVA tables indicates no statistically significant interaction effect between the two independent variables of the study. This lack of significant interaction effect, which is graphically displayed for the open-ended test below, implies that the

main effects of sex and personality type are in a straightforward relationship and the effects of one independent variable do not differ at various levels of the other one.

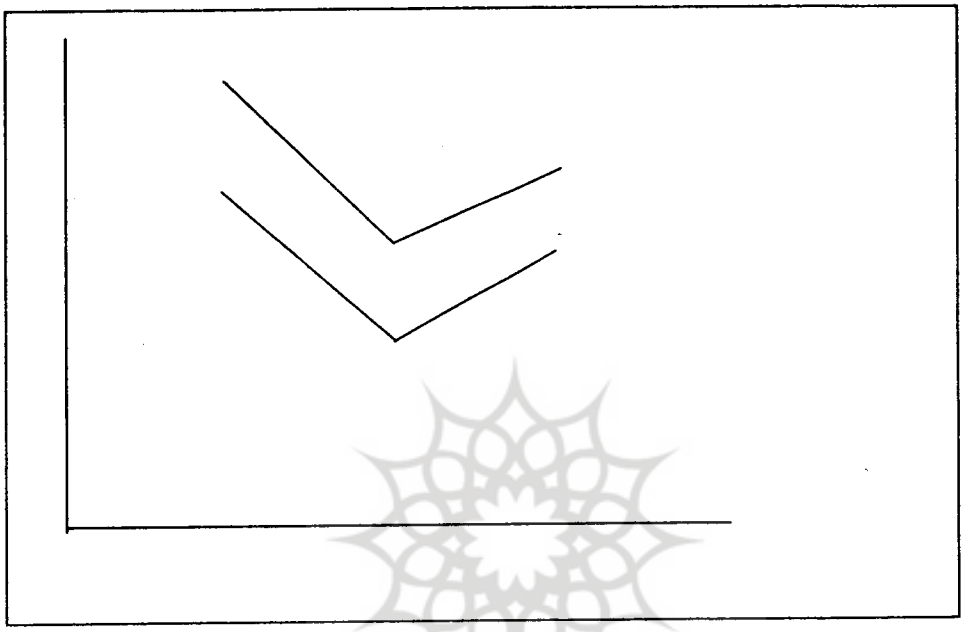

Figure 1 Interaction effect between male and female subjects on the open-ended test

IV. Pedagogical Implications

Based on our findings, the measurement ability and accuracy of a more practical test method like multiple-choice does not seem to be different from other time-consuming and less reliable methods such as open-ended. Thus, language teachers and testers can be recommended not to bother themselves with utilizing the latter group of test methods and instead make every endeavor to develop sounder multiple-choice tests with a higher degree of

reliability which can, in turn, satisfy the necessary conditions for their validity too.

Our findings show that contrary to test method, personality type and gender have strong impact on test performance. Therefore, they need to be considered by language testing theorists in taking every step in quest of test validity and specially in constructing test validity frameworks.

Impact of personality type and gender on test performance need to be considered by language teachers and testers too. What they need to be scrupulous about is the assumption that simply because of being of a particular personality type or being a male or female, a student may succeed in learning a foreign language or the outcome may be the reverse.

Our conclusion pointing to the extroverts' higher rate of success in EFL learning because of their outgoingness and interaction-orientedness may imply that EFL teachers need to create classroom environments where their students can receive the maximum degree of language practice by getting involved in meaningful communicative activities. It may also imply that they need to promote those learning strategies which tap such traits.

On the basis of the same finding, we can recommend the syllabus designers and materials developers, to the extent possible, to try to have reflections of the above-mentioned traits in language teaching materials. For example, they can

design learning activities and tasks which tap the above-mentioned traits.

Finally, the strong positive effect of femininity and extroversion on EFL learners may imply that the university admissions policy-makers in our country need to reconsider their policies and criteria with respect to those students who are admitted to the university EFL programs. It seems that female extroverts and male extroverts have got the first and second priorities of admission into such programs respectively.

V. References

Abbasy A. (1995). *Extrqversion/Introversion and Success in English Language Reading Comprehension.* Unpublished M.A. Thesis, Allameh TabaTabaee University, Tehran.

Alderson, J. C. (1991). "Language testing in the 1990s: How far have we got? How much further have we to go?". In S. Anivan (ed.), Singapore: SEAMEO Regional Language Center.

Alderson, J. C. and Banerjee, J. (2001). "Language testing and assessment (Part I)". *Language Testing 34, 213 - 236.*

Astika, G., Carrell, P., and Moneta, Prince (1996). "Personality types and Language learning in an EFL context." *Language Learning* 46, 75 - 99.

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

Bachman, L. F. (1991). "What does language testing have to offer". *TESOL Quarterly* 25, 671 - 704.

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

Boyle, J. P. (1987). "Six differences in listening vocabulary". *Language Learning* 37, 273 – 284.

Brown, H. D. (1973). "Affective variables in second language acquisition". *Language Learning* 23, 231 - 244.

Burger, J. M. (1990). *Personality*. Belmont, California: Wadsworth Publishing Company.

Busch, D. (1982). "Introversion-extroversion and the EFL proficiency of Japanese students." *Language Learning* 32, 109 - 132.

Chastain, K. (1975). "Affective and ability factors in second language acquisition". *Language Learning* 25, 153 - 161.

Dewacle, J. (1993). "Variation in the morphosyntactic and lexical systems of French-based interlanguages", in Kettemann, B. and Wieden, W. (eds.). Gunter Narr Verlang: Tubingen.

Eisenstein, M. (1982). "A study of social variation in adult second language acquisition". *Language Learning* 32, 367-391.

Elder, C. (1997). "What dose test bias have to do with fairness". *Language Testing 14, 261 - 275.*

Eysenck, H. J. (1952). *The Scientific Study of Personality.* New York: Routledge and Kegan Paul.

Eysenck, H. J. and Eysenck, S. B. G. (1991). *Manual* of *the Eysenk Personality Scales (EPS Adult).* London: Hodder & Stoughton.

Farhady, H. (1982b). "Measures of language proficiency from the learner's perspective". *TESOL Quarterly* 16, 43 - 59.

Gibson, H. B. (1981). *Hans Eysonck: The Man and his Work.* London: Peter Owen.

Gross, D. (1983). "Sex differences in achievement". *System* 11, 159-162.

Guiora, A. ; Acton, W. ; Erard, R. and Strickland, F.(1980). "The effects of Benzodiazepine(Valium) on permeability of language ego boundaries". *Language Learning* 30, 351 - 363.

Hesamy, A. (1994). *Test Method, Test Taker's Cognitive Style (Field Dependence-Independence) and their Effect on Test Performance.* Unpublished M.A. Thesis. Allameh Tabatabee University, Tehran.

Howe, K. (1995). *//A: HOWEVA ~ 1. HTM,* Online.

Larsen-Freeman, D. and Long, M. H.(1993). *An Introduction to Second Language Acquisition Research.* London: Longman.

MacIntyre, P. and Charos, C. (1996). "Personality, attitudes and affect as predictors of second language communication." *Journal of Language and Social Psychology 15, 3-26.*

Messick, S. A. (1988b). "Validity" in Linn 1988(ed.), third edition: 13 – 103.

Messick, S. (1988a). "The once and future issues of validity: Assessing the meaning and consequences of

measurement". In H. Wainer & H. I. Brown(eds.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Messick, S. (1996). "Validity and washback in language testing". *Language Testing* 13, 241 - 256.

Naiman, N.; Frohlich, M.; Stern, H.; and Todesco, A. (1978). "The good language learner". *Research in Education Series, No.* 7. Toronto, Ontario Institute for Studies in Education.

Shohamy, E. (1984a). "Does the testing method make a difference? The case of reading comprehension". *Language Testing* 1, 147 - 180.

Shohamy, E. (1997). " Testing methods, testing consequences: Are they ethical? Are they fair?". *Language Testing* 14, 340 - 349.

Shohamy, E. and Stansfield, C.(1991). " The Hebrew oral test: An example of international cooperation". *AILA Bulletin 7.*

Shohamy, E.; Reves, T. and Bejerano, Y. (1986). " Introducing a new comprehensive test of oral proficiency". *English Language Teaching Journal* 40, 212 - 220.

Shohmy, E.; Donitze-Schmidt, S. and Waizer, R. (1993). "The Effect of the Elicitation Method on the Language Samples Obtained on Oral Tests". Paper presented at the annual Language Testing Colloquium , Cambridge, UK.

Spurling, S. and Illyin, D. (1985). "The impact of learner variable on language test performance". *TESOL Quarterly* 19, 283 - 301.

Stansfield, C. and Kenyon, D. (1988). "Development of the Portuguese Speaking Test". Year one report on development of semi-direct tests of oral proficiency in Hausa, Hebrew, Indonesian and Portuguese. Alexandria, VA: ERIC Document Reproduction Service ED 296 - 586.

Suter, R. (1976). "Predictors of pronunciation accuracy in second language learning." *Language Learning:* 26, 233 – 253.

Swain, M. and Burnaby, B. (1976). "Personality characteristics and second language learning in young children: A pilot study". *Working Papers on Bilingualism* 11, 76 - 90.

Takala, S. and Kaftandjirva, F. (2000). "Test fairness: a DIF analysis of an L2 vocabulary test". *Language Testing* 17, 323 - 340.

Valizadeh, Gh. (1999). Effect of Gender on Iranian EFL Learners' Pronunciation. Unpublished M.A. Thesis, Allameh Tabatabee University, Tehran.