Journal of Language and Linguistic Studies
Vol.6, No.2, October 2010

The Impact Analysis of Psychological Reliability of Population Pilot Study For Selection of Particular Reliable Multi-Choice Item Test in Foreign Language Research Work

Seyed Hossein Fazeli

fazeli78@yahoo.com

Abstract

The purpose of research described in the current study is the psychological reliability, its' importance, application, and more to investigate on the impact analysis of psychological reliability of population pilot study for selection of particular reliable multi-choice item test in foreign language research work. The population for subject recruitment was all under graduated students from second semester at large university in Iran (both male and female) that study English as a compulsory paper. In Iran, English is taught as a foreign language.

Key word: psychological reliability, reliability, testing, pilot study

1. Introduction

A fundamental concern in development and use of language tests is identification of potential sources of error. We must concentrate on such errors because it is clear that test performance is affected by factors other than the abilities we want to measure (Bachman, 1995). In this way no matter what type of test, no matter what length, and no matter what modality it has, any task referred to as a test must have certain characteristics to avoid some errors which affect the results of the test. One of these basic characteristics of test, that is absolute crucial, is reliability (Hatch & Farhady, 2007, Jafarpaur, Farhady

& Brijandy, 20008; Jafarpuor, 1999; Farhady, 2008). Reliability of a test is defined as "the extent to which the result can be considered consistent or stable" (Brown, 1998, p. 98), similarly, it is defined as "a quality of test scores which refers to the consistency of measures across different times, test forms, raters and other characteristics of the measurement content" (Mousavi, 1999, p. 323) , or in its' simplest definition refers to "consistency often meaning instrument consistency" (Mackey & Gass, 2005, p. 128).

In language testing, when we study ways to estimate reliability, we would find methods such as test-retest, parallel tests or equivalent forms, internal consistency methods (split –half, Kuder-Richardson formula 20, and Kuder-Richardson 21), Scrorer reliability, and standard error of measurement (Brown, 1998; Mousavi, 1999; Riazi, 1998; Jafarpaur, 1995; Bachman, 1995) which are based on the persons' true score (which is constant). Persons' true score is defined as "the average of observed scored over infinite number or parallel test assuming that the person is in a steady state" (Hatch & Farhady, 2007, p. 246). Clearly the concept of methods to estimate reliability shows that reliability is based on the statistical state of test items only , and statistically no other more factors interfere quantity of test reliability, however personal attributes and the random factors alongside test method affect language test scores (Bachman, 1995).

Alongside the investigation of reliability for both of logical and empirical research aspects, we must identify sources of error and estimate the magnitude of their effects on test scores (Monsavi, 1999). Identification of such source of errors needs to distinguish the effect of language abilities that we want to measure separately and without interference from effect of other factors which is a particularly complex problem and it is fluctuation in the psychological conditions of the learners that can effect up to great extent which was not considered significantly during the language testing study. It is important that researchers report about psychological factors statistically, regardless type of reliability of particular test and even before estimate reliability of particular test. Such report of estimating of psychological reliability, it should be base of the reliability of test that is discussed widely today. This report here is called as psychological reliability. Psychological reliability is different from reliability of test in:

a) It will be varied from particular population to another regardless structure of test or any change in structure of test.

b) Psychological reliability determines certain characteristics of particular population of the test takers.

c) Psychological reliability is the base for test reliability; in other words, it shows that really reliability of test is accepted or not. To claim about reliability of test, we should estimate the psychological reliability of population statistically, otherwise we cannot claim about type of reliability correctly. In fact, giving test in order to estimate reliability of test needs to estimate psychological reliability of population statistically, and if the result was acceptable, we continue the procedure to estimate reliability of test.

In the current study, the investigator aims to test different samples to show that we cannot claim a bout reliability of test directly and it needs some other basic reliability of population which is called psychological reliability.

2. Methodology

2.1 Participants

The population for subject recruitment was all undergraduate students from second semester at a large University in Iran for academic year 2008-2009, which were volunteered to participate in this study. The participants were both male and female that were told this study would not affect the final results of their course. Based on the English language proficiency test, they were divided into two groups that were same number of member (every group forty and two students). The first group includes the students who are advanced students in English language proficiency(22 male and 20 female), that is called as a “High-Level Group” and the second group includes the students who are elementary students in English language proficiency(23male and 19

female) that is called as an “Elementary-Level Group” here. The range of all participants’ age was varied from 20 to 30 years old, and the mean of their age was 24.2.

2.2 Instrumentation

The following instruments were used in the current study:

a) Questionnaire elicited information regarding demographic profile of the respondents (e.g. age),

b) Materials which are developed as a test, includes fifty multi-choice questions. Every question has four options that only one of them as a correct answer,

c) Michigan Test of English Language Proficiency (MTELP) 2007, in order to have two advance and elementary language proficiency groups in English in the current study.

2.3 Procedure

2.3.1 Questionnaire

The questionnaires were distributed two weeks before testing procedure in order to elicit information regarding demographic profile of the respondents. The students completed answering the questionnaires in the class and returned the forms to the lecture. The needed details regarding questionnaire, were explained by the lecture.

2.3.2 Proficiency Test

The Michigan Test of English language proficiency was used to determine the level of the students’ English proficiency one week before the treatment. The mentioned proficiency test was studied in pilot study to find out its reliability that it was estimated 6.8.

2.3.3 Testing procedure

In the current study, before the beginning of taking test, two groups were told when they are sure about the correct answer; they select that correct choice, otherwise do not select any option. It was told them about the importance of the results of this test that it will be applied somewhere, and the results of this study are as reference for decision to apply for lager population.

Before the beginning of taking test, the High-Level Group was asked when they select any choice for every question as correct answer; they should mention the reason of selection of that choice, otherwise their answer is not scored. For Elementary-Level Group, before the beginning of taking test, only they were asked to select the correct answer and they did not tell to mention reason for their selection of choice for every question as correct answer. After finishing test, the question papers were collected. The second step of procedure of the Elementary-Level Group is that the same former questions of test were given to the students again. But they were asked to mention the reason for selection of the options as was done for the High-Level Group procedure. Enough time as had been estimated in pilot study before, was given to the two groups regarding their activities of testing.

3. Data Analysis

3.1 Scoring procedure

Testing procedures of two groups are concluded in five types of scores:

1. Numbers of choices as answer were selected by the High-Level Group, whether correct or wrong, which are called as First Scores.

2. Numbers of correct answers that were selected by the High-Level Group, which are called as True Scores.

3. Numbers of choices as answer were selected by the Elementary-Level Group, whether correct or wrong, which are called as First Scores.

4. Numbers of choices as answer were selected by the Elementary-Level Group, whether correct or wrong, which are called Second Scores. Second Scores, are obtained after that the Elementary-Level Group was told to mention reason for selection of choice as answer.

5. Numbers of correct answer that were selected by the Elementary-Level Group, which are called as True Score.

3.2 Data analysis procedure

In this study, analysis of obtained data was performed using the SPSS software version 16.

4. Results and Findings

Firstly let consider the tables of Elementary-Level Group separately.

Table one

	Mean	N	Std. Deviation	Std. Error Mean
First Score	42.9048	42	6.92049	1.06785
Second Score	13.1190	42	4.03764	.62302

In table one, the means of First and Second Scores of the Elementary-Level Group are shown to make some clearness for the following table and results.

Table two

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
First Score - Second Score	29.7857	6.27630	.96845	27.82988	31.74155	30.756	41	.000

In table two, the mean difference between the means of First and Second Scores of the Elementary-Level Group is shown to focus and emphasize on big significant difference which can be useful to show the un-reliable condition of population of the Elementary-Level Group.

Table three

	Mean	N	Std. Deviation	Std. Error Mean
First Score	42.9048	42	6.92049	1.06785
True Score	.9286	42	1.71639	.26484

In table three, the means of First and True Scores of the Elementary-Level Group are shown to make some clearness for the following table and results.

Table four

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
First Score - True Score	41.9762	6.97902	1.07689	39.80137	44.15101	38.979	41	.000

In table four, the mean difference between means of First and True Scores of the Elementary-Level Group is shown to focus and emphasize on big significant difference which can be useful to show the un-reliable condition of population of the Elementary-Level Group.

Table five

	Mean	N	Std. Deviation	Std.Error Mean
Second Score	13.1190	42	4.03764	.62302
True Score	.9286	42	1.71639	.26484

In table five, the means of Second and True Scores of the Elementary-Level Group are shown to make some clearness for the following table and results.

Table six

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Second Score - True Score	12.1904	3.59022	.55398	11.07169	13.30927	22.005	41	.000

In table six, the mean difference between the means of Second and True Scores of the Elementary-Level Group is shown to focus and emphasize on the affect of suggestion to write the reasons of the options which were selected among population of the Elementary-Level Group. However the mean difference is significant.

Secondly let consider the table of High-Level Group separately.

Table seven

	Mean	N	Std. Deviation	Std. Error Mean
First Score	27.2857	42	10.83870	1.67245
True Score	13.9762	42	7.27337	1.12230

In table seven, the means of First and True Scores of the High-Level Group are shown to make some clearness for the following table and results.

Table eight

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
First Score - True Score	13.3095	6.16239	.95088	11.38919	15.22986	13.997	41	.000

In table six, the mean difference between the means of First and True Scores of the High-Level Group is shown to focus and emphasize on the affect of suggestion to write the reasons of the options which were selected among population of the High-Level Group. However the mean difference is significant. Moreover the English language proficiency level of this group is more than the Elementary-Level Group.

Thirdly let consider the tables of Elementary Level Group and High Level Group in contrast.

Table nine

	Both Groups	N	Mean	Std. Deviation	Std. Error Mean
First Score	Elementary Level Group	42	42.9048	6.92049	1.06785
	High Level Group	42	27.2857	10.83870	1.67245
Second Score	Elementary Level Group	42	13.1190	4.03764	.62302
	High Level Group	42	99.0000	.00000	.00000
True Score	Elementary Level Group	42	.9286	1.71639	.26484
	High Level Group	42	13.9762	7.27337	1.12230

In table nine, the First, Second and True scores of the Elementary-Level Group, and the First, and True scores of the High-Level Group are shown to conclude and make general view on the results of the both groups.

5. Discussion

Let calculate different scores that were obtained from the two groups.

(Elementary-Level Group = ELG), (High-Level Group = HLG)

The Elementary-Level Group

First Score Mean: 24.9048 = % 85.8096

Second Score Mean: 13.1190 = % 26.2380

True Score Mean: .9286 = % 1.8572

↓↓↓

(First Score Mean) – (Second Score Mean) = 29.7858 = % 59.5716

(First Score Mean) – (True Score Mean) = 41.9762 = % 83.9524

(Second Score Mean) – (True Score Mean) = 12.1904 = % 24.3808

The High-Level Group

First Score Mean: 27.2857 = % 54.5714

True Score Mean: 13.9762 = % 27.9524

↓↓↓

(First Score Mean) – (True Score Mean) = 13.3095 = % 26.6190

Regarding both the Elementary-Level Group (ELG) and the High-Level Group (HLG)

(First Score Mean of ELG) – (First Score Mean of HLG) = 15.6191 = % 31.2382

(True Score Mean of HLG) – (True Score Mean of ELG) = 13.0476 = % 26.0952

(First Score Mean of HLG) – (Second Score Mean of ELG) = 14.1667 = % 28.3334

[(First Score Mean of ELG) – (First Score Mean of HLG)] – [(True Score Mean of ELG) – (True Score Mean of HLG)] = 2.5715 = % 5.1430

Alongside the control of many factors which may interfere in the procedure of current work, logically analysis the above calculation leads to understand that:

a) The amount %59.5716 Mean Differences between the First Score and Second Score of answering procedure of the Elementary-Level Group, show that the students of this group are %59.5716 far from the point to consider the testing procedure as real procedure. This farness (distance) is as un-reliable psychological amount of the Elementary-Level Group.

b) The amount % 26.2380 as the Second Score of the Elementary-Level Group can be as psychological reliable amount for the Elementary-Level Group here. For every group, should be some maximum level of Mean that up to that level shows

the psychological reliability of particular group. Regarding the above calculation, for the Elementary-Level Group of the current study, approximately %3o Mean is the maximum level of answering of this group .

c) The amount .9286 as the True Score of the Elementary-Level Group is as true ability of this group.

d) In procedure of the Elementary-Level Group, the contrast of amount of percents between ((First Score)- (True Score) = %83.9524) and ((Second Score) – (True Score) = %24.3808), show that later one (%24.3808) is more reasonable than the former (%83.9524) in multi-choice test when for every question there is four choices and one is correct and the others are wrong (again, it is mentioned that in the current study, multi-choice test is applied and every question has four choices that one is correct and the others are wrong)and since every student for any question is searching for one correct option that is %25.

e) Regarding the First Score of the High-Level Group (which was asking to mention reason for choice), the First Score of the Elementary-Level Group (which not asking to mention reason for choice), more English language ability of the students in the High-Level Group for amount of correct answering compare to the Elementary-Level Group (%26.0952 Mean difference) and their higher proficiency of English language , and the Second Score of the Elementary-Level Group, %31.2382 Mean difference between two groups are because of applying the specific strategy to whether asking or NOT asking to write reason of choice the options. Such strategy is as an instrument to decrease the percent of psychological un-reliability of particular group.

f) Much of distance among Mean of the First Scores and True Scores of the both groups can show amount of difference regarding psychological reliability between two groups, which it is estimated as %5.1430.

High Level Group: HLG, Elementary Level Group: ELG

[(First Score of ELG) – (First Score HLG)]- [(True Score of ELG) - (True Score of HLG)] = 2.5715 = % 5.1430

6. Conclusion

The current study suggests that psychological reliability refers as a certain characteristic of the sample of the test takers which is basic factor in decision to select particular sample of the learners as acceptable sample; otherwise it is impossible to claim that particular test has acceptable reliability.

The investigator aims to test that the procedure of reliability checking has two steps which are as first step, checking psychological reliability of the samples of learners in order to determine reliable sample of test takers, and as second step checking reliability of test itself. Every one of the two steps, is necessary, but is not enough without another to claim whether particular test is applicable and reliable or not.

The importance of application of psychological reliability in one pilot study when one investigator aims to test that particular test is reliable for the larger population. Because of such decision, having of valid and real results of pilot study of particular test of particular sample as pilot sample of test takers in the research work, is more focused (Brown, J.D. & Rodgers, T.S, 2004). Trust on psychological reliability of the rest takers and reliability of test together causes that the investigator trusts on the particular reliable characteristics of test.

Further research is need to better determine the strength of association among different proficiency levels, other types of test, other skills of language, other state of English as a second language instead of foreign, particular gender ,other ages, and so on. A future agenda for psychological reliability researchers should specially and thoroughly address to the particular formulae to be applied to assess psychological reliability better and help the research that better sample selection will take place.

Application for language testing from this study is that psychological reliability plays the most necessary role in determination of reliable sample of test takers, and in language testing; the researcher should select psychological reliable samples and then try to do checking reliability of test. The most important application of psychological reliability in pilot study because of the obtained results will be applied for larger population; therefore the type of such results (weather right or wrong) can affect the results of whole population significantly.

Although this study, clearly contributes to our understanding of psychological reliability but there are limitations to consider. Firstly, the focus of the study was on language testing for statistical significance. Secondly, future research should consider mixed design or studies for that examine qualitative aspects of the topic. Several of the limitations to this study are ones common in the literature, the needs for a large n-size; need to conduct similar experiments with different population and proficiency levels and so on.

References

Bachman, L.F. (1990/1995). *Fundamental Considerations in Language Testing*. Oxford: Oxford university press.

Brown, J.D. (1988/1995). *Understanding Research In Second Language learning*. Cambridge: Cambridge University Press.

Brown J.D & Rodgers. T.S .(2002). *Doing Second Language Research*. Oxford: Oxford university press.

Farhady, H., Jafarpur, A. & Brijandi, P. (2000). *Testing Language skills: From Theory to practice*. Tehran: SAMT Publication.

Farhady, H. (2008). *Research Method in Applied Linguistics*. Tehran: Pyame Noor University press.

Hatch, E. & Farhady H. (2007) *Research Design and Statistics*. Tehran : Rahnama Publication.

Hughes, A. (1989/2003). *Testing for Language Teachers*. Cambridge: Cambridge university press.

Jafarpur, A. (1997). *An Introduction to Language Testing*. Shiraz: Shiraz University Press.

Mackey, A. & Gass, S.M. (2005). *Second Language Research*. New Jersey: Lawrence Erlbaum Associates publishers.

McNamara, T. (2000) *Language Testing*. Oxford: Oxford University Press.

Mousavi, S.A. (1999). *A Dictionary of Language Testing*. Tehran: Rahnama Publication.

Riazi, A.M. (1999). *A Dictionary of Research Methods: Quantitative and Qualitative*. Tehran: Rahnama publication.

Seyed Hossein Fazeli is pursuing his Ph.D. program in Linguistics regarding English Language Learning Strategies at University of Mysore, Mysore, India. His interests include Language Learning Strategies, and analysis of approaches regarding Language Learning Strategies. He can be contacted at fazeli78@yahoo.com or fazeli78@hotmail.com, Mobile Number: 0091-8095042070(0091-9632329606).