

## The construct validation of Iranian student's Reading Comprehension Tests through Multitrait-Multimethod Procedure

Sohrab Rezaei

Assistand Professor Allameh Tabatabaei University  
[Sohrab\\_rezaei@yahoo.com](mailto:Sohrab_rezaei@yahoo.com)

**Abstract:** Construct validity is a process of investigating what a test measures, in which one validates a test not against a criterion or another test, but against a theory. There is a specialized construct validation procedure called Multitrait-Multimethod procedure. This study has aimed at determining the impact of testing method on Iranian EFL learner's reading comprehension using multitrait-multimethod procedure. In other words, this study has been an attempt to determine whether the learners who take multiple choice cloze tests on vocabulary and grammar would score significantly higher than those who take multiple choice paraphrase and multiple choice comprehension tests on vocabulary and grammar or vice versa. A population of 100 grade 4 male undergraduate students majoring English translation in Karaj Azad university have been selected as subjects. Following administration of a TOFEL test 35 students have been selected as final subjects to be tested. The research procedure has included three methods namely comprehension, cloze and paraphrase, each of which contained two traits of vocabulary and grammar. A battery of statistical analysis namely correlational analysis, multivariant analysis of variances and factor analysis has been employed to investigate the results. The findings have revealed that there has been a strong correlation among methods. They have also revealed that the method, skill and the interaction between them have a high impact on the mean scores of different tests. The findings have also indicated that it has been due to the teachers and test practitioners to make use of different methods in their testing methodologies.

[Sohrab Rezaei. **The construct validation of Iranian student's Reading Comprehension Tests through Multitrait-Multimethod Procedure.** *Life Sci J* 2012;9(3):2118-2126] (ISSN:1097-8135).  
<http://www.lifesciencesite.com>. 306

**Key words:** validity, testing, construct validation, multitrait-multimethod, comprehension, cloze, reading, paraphrase.

### Introduction

Validity is one of the important concepts in testing. A valid test is one that actually tests what the designer of the test intends to test. Regarding the term "valid test" in relation to testing, there are a number of cautions to be born in mind. Validity deals with the results of a test not with the instrument itself. In this respect, according to Gronlund (1976), we speak of the validity of the test results, or more specifically, of the interpretation to be made from the results. Validity is a matter of degree. Therefore, we should avoid thinking of a test results as valid or not valid. Validity is best considered in terms of categories that specify degrees, such as a high validity, moderate validity, and low validity. Validity is not general but specific. Accordingly, if a test of vocabulary measures vocabulary and nothing else, it is a valid test of vocabulary. In this regard Ingram (1977), states: "when a test measures that which is supposed to measure, and nothing else, it is valid." (P.18)

One of the most important ways of evaluating validity is construct validity. Construct validity is a process of investigating what a test measures. Through which one validates a test not against a criterion or another test, but against a theory. There is a specialized construct validation procedure called the

multitrait-multimethod convergent-divergent procedure. The procedure was first described by Campbell and Fiske (1959) and was first recommended for use in the evaluation of language proficiency measures by Stevenson (1974). It is based on the assumption that a test score is a function both of the trait the test measures and of the method by which it is measured. In order to measure the relative contributions of trait and method, it is necessary, for statistical reasons, that two or more traits each be measured by two or more distinct methods. It is for this reason that the procedure is called multitrait-multimethod (MTMM) procedure.

Campbell and Fiske (1959) argue that the demonstration of construct validity requires both convergent validity and discriminant validity, that is, multiple of the same construct should be substantially correlated with each other, but less correlated with indicators of other constructs. They propose collecting measures of more than one trait, each of which is assessed by more than one method. Convergent validity is inferred from agreement between measures of the same trait assessed by different methods. Discriminant or divergent validity refers to the distinctiveness of the different traits, and is inferred from the relative lack of correlation between different

traits. Support for these characteristics is based upon inspection or analysis of a multitrait- multmethod (MTMM) matrix. MTMM matrices have been analyzed by a variety of different procedures. The most frequently employed procedures have been the original Campbell- Fiske Criteria, and an ANOVA model. More recently, applications of confirmatory factor analysis have been applied to MTMM matrices. Though the approach has been described under a variety of different labels: restricted factor analysis (Brounch & Wolins, 1970), confirmatory factor analysis (Kenny, 1976, Werts, Joreskog, Linn, 1972), path analysis (Schmitt, 1978; Schmitt et al., 1977), and exploratory factor analysis (Lomax and Algina, 1979).

### **Review of Related Literature**

Validity is a frequently misunderstood concept. It is often erroneously believed that a test is simply valid or not valid as if validity were a property of the test itself. In fact, as Cronbach (1971) has pointed out, one does not validate a test. One validates "an interpretation of data arising from a specified procedure. The elements affecting validity include, among others, the test itself, the setting in which the test is administered, the characteristics of the inferences intended to be drawn from the test. The general purpose of the validation procedure is to investigate the extent to which inferences can properly be drawn from performance. The process of collecting evidence of the extent to which such inferences are warranted is called validation.

Although validity has traditionally been discussed in terms of different types, psychometricians have increasingly come to view it as a single, unitary concept. Messik (1988) proposes a unified framework of validity that brings consideration of value implications and social consequences into the validity framework. (P 20). Messik describes validity as an internal evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores. (P13). One of the first characteristics of a test that we examine is its content. If we cannot examine an actual copy of the test we would generally like to see a table of specification and example items, or at least a listing of the content areas covered, and the number of, or relative importance of each area. The consideration of test content is an important part of both test development and test use. Demonstrating that a test is relevant to and covers a given area of content or ability is therefore a necessary part of validation. There are two aspects of this part of validation; content relevance and content coverage. (Messick, 1988). Henning (1987) refers to response validity to describe the extend to which examinees responded in

the manner expected by the test developers. If examinees respond in a haphazard or non-reflective manner, their obtained scores may not represent their actual ability. Also if instructions are unclear and the test format is unfamiliar to the students, their response may not reflect true ability and the test may be said to lack response validity. This type of validity investigates the correspondence between the scores obtained from the newly developed test and the scores obtained from some independent outside criteria (Farhady et al, 1994). There are two types of criterion validity; concurrent and predictive. In concurrent validity the criterion behavior may be concurrent with, or occur nearly simultaneously with administration of the test (Backman, 1990). Information on concurrent criterion relatedness is undoubtedly the most commonly used in language testing. Such information typically takes one of the two forms:(1) examining differences in test performance among groups or individuals at different levels of language ability or (2) examining correlations among various measures of a given ability. This type of validity is used to determine how well test scores predict some future behavior. Examining predictive utility is often problematic because, the criterion behavior that we want to is often a complex one that may depend upon a large number of factors in addition to language ability.

### **Construct validity**

Cronbach and Meeld (1965), define a construct as a postulated attribute of people assumed to be reflected in test performance. Thus construct can be viewed as definition of abilities that permit us to state specific hypotheses about how these abilities are or are not related to other abilities, and observed behavior. The notion of construct validity was formerly originated in 1954 with the publication of technical recommendations of a committee of the American Psychological Association. While construct validity is empirical in nature because it involves the gathering of data and the testing of hypotheses, unlike concurrent and predictive validity, it does not have any one particular validity coefficient associated with it. The purpose of construct validation is to provide evidence that underlying theoretical construct begins with a psychological construct that is part of a formal theory. Henning (1987) argues that we have a very great difficulty in establishing construct validity. It is the fact that the construct itself cannot be measured directly. Backman (1990) argues that construct validity concerns the extend to which performance on test is consistent with predictions that we make on the basis of a theory of abilities or constructs. Messik (1980) sees construct validation as a unifying concept that investigates criterion and content considerations into a common framework for testing rational hypotheses about theoretically relevant relationships.

There are different ways through which one can establish the construct validity of a given test. One of the versions is the multitrait –multimethod (MT\_MM), convergent divergent validity developed by Campbell and Fiske (1959) and its modified form used by a number of experts. They proposed the use of multimethod matrix in which inter-correlations among several traits each measured by several methods are appraised for evidence of discriminant and convergent validity. Four informal criteria have been suggested by Campbell and Fiske for the purpose of evaluating the MT\_MM matrix. First the correlations between similar traits measured by different methods. Second, the convergent validities should be higher than the correlations between different traits measured by different methods. Third, the convergent validities should be higher than the correlations between different traits measured by the same method. Finally, a similar pattern of trait intercorrelations should be apparent in the hetrotrait-monomethod sub matrices and hetrotrait- hetromethod sub matrices. This criterion was questioned by many scholars for example, Jackson (1967), Conger (1971), Tucker (1976) and Seidman (1974). Alwin (1974), Joreskog (1989), Kalleberg and Kluegel (1975), and Werts and Linn (1990), among others, have proposed a path analysis approach for evaluating MT\_MM matrices. Jackson (1969) proposed multimethod factor analysis which involves replacing the monomethod blocks of the matrix with identity matrices, thus removing method variance and basing estimated scores on the portion of variance common to more than one method of measurement. Marsh and Hocevar (1983) proposed confirmatory factor analysis as superior method to use in the analysis of multitrait – multimethod data.

Jackson and Singer (1996) got to the conclusion claiming the correlation between males and females upon four different traits are higher in the hetro-method blocks for bearing the same name as compared with the values for the heterotrait-heteromethod correlations. Backman and Palmer (1985) in their study on construct validity of speaking and reading ability concluded that the effect of test method on test score is strong and meaningful. Farr and Jongsma (1993) in their study examined convergent and discriminate validity of three components of a set of integrated reading/writing assessments. The three factors assessed by the test included Response to Reading, Management of content and command of Language. Tepper et al (1992), tried to establish the discriminant and convergent validity of the Problem Solving Style Questionnaire (PSSQ). In this investigation they tried to assess the validity of the PSSQ's two subscales (abstractness/ concreteness and action/ reflection).

Findings of the research provided further support for construct validity of new scales. Hierarchically nested confirmatory factor analysis demonstrated the superiority of a model representing the latent variables abstractness/ concreteness and reflection /action over a one trait model (providing evidence of discriminant validity) and a null model (providing evidence of convergent validity).

#### **Statement of the problem**

Regardless of the recent researches and findings in the field of teaching reading skill to foreign /second language learners, many teachers are still teaching via different methods. Following the methods, teachers try to use different kinds of testing methods to measure the students reading comprehension ability.

The present study has aimed at determining the impact of testing procedures on reading comprehension ability of the Iranian EFL learners. In other words, this study has been an attempt to determine whether the learners who take a multiple choice cloze test would score significantly higher on reading comprehension than those who take multiple choice comprehension tests, or vice versa.

Nowadays most teachers have a great tendency to use the cloze test technique to check learner's ability and progress on reading comprehension. Therefore, if cloze technique is theoretically a good method to test reading comprehension why do teachers and other examiners use other methods such as paraphrase and multiple choice items in practice?

Concerning the problems, mentioned above, the researchers set forth the following objectives in designing this study:

- 1- To investigate and compare the impact of cloze test vs. paraphrase tests and Comprehension tests on reading comprehension ability of the subjects under study to find out if there is any correlation among them.
- 2- To investigate and compare the impact of cloze test vs. paraphrase test procedure on reading comprehension ability of the subjects under study and find out if there is significant correlation between them.
- 3- To investigate and compare the impact of paraphrase tests vs. comprehension tests on reading comprehension ability of the subjects under study and to fine out if there is any significant correlation between them.
- 4- To investigate and compare the impact of cloze test vs. comprehension test on reading comprehension ability of the subjects under study to find out any significant correlation between them.

#### **Research Questions**

To come up with logical answers, the following research questions have been asked in order to

determine the construct validity of reading comprehension tests through traits of vocabulary and Grammar and methods of multiple choices cloze, MC paraphrase and MC comprehension called multitrait-multimethod (MTMM) procedure.

- IS there any significant correlation among the MC cloze, MC paraphrase and MC comprehension tests on Vocabulary?

- IS there any significant correlation among the MC cloze, MC paraphrase, and MC comprehension tests on grammar?

- IS there any significant correlation among the MC cloze, MC paraphrase, and MC comprehension tests on vocabulary and grammar?

- IS there any underlying construct for the MC cloze, MC paraphrase and MC comprehension tests on Vocabulary and grammar?

### **Significance of the Study:**

Language testing is central to language teaching. It provides goals for language teaching, and it monitors, for both teachers and learners' success in reaching those goals. Its influence on teaching is strong. It provides methodology for experiment and investigation in language teaching and language learning/ acquisition.

The findings of this research would have both theoretical and practical implications for the field of language testing. Obviously, through these findings one positive step would be set forward to make decision upon education careers of testing English as a foreign/ second language. This study has direct relevance to developing test methods as well as moderating methodology and text- books used in language instruction. In other words, the result of such a research would have been of great interest to most language testing practitioners interested in improving the quality of their testing methodologies. Materials developers and curriculum designers may also find the findings of this study useful for instructional goals.

To be more specific, in most language testing situations especially in the case of our country, cloze test technique and multiple choice items have been given more weight in testing reading comprehension ability of learners. More over, nowadays there have been a lot of tests with different methodologies, constructed for beginners and advanced levels, available in the market, whereas the validity of the

tests and their methodologies have rarely been verified, and a lot of money and time is spent on producing them. Therefore, investigation in this regard is highly warranted.

### **Limitations of the study:**

1- In this study only two traits (i.e. Vocabulary and grammar) have been tested through three different procedures (ie. MC cloze, MC paraphrase and MC comprehension).

2- Due to the nature of the study, there has been a need to have different kinds of tests, and hence there has been no single test for the purpose of the study, the researchers have had to use different valid tests to collect data.

### **Method and procedure**

To pave the way for the application of this study based on the questions mentioned, several steps have been required in the acts of subject selection, preparation and selection of the needed tests and application of the selected tests to the subjects as well as the interpretation of the results.

### **Subjects:**

A sample of one hundred male students from Karaj Islamic Azad University has been chosen randomly. The subjects were grade 4 under-graduate students majoring in English translation. To ensure the homogeneity of the subjects a TOFEL test was administered. The subjects have been ranked according to the score through the application of the t-Test.

Finally thirty five students who clustered around the mean score have been selected for the study. The rationale behind choosing the subjects from this level and field of study was that the nature of given tests needed subjects who should have been familiar with different kinds of testing procedures namely cloze, comprehension and paraphrase.

### **Procedure:**

After selecting the required subjects, another test has been administered to measure the relative contribution of traits and methods. Two traits of "grammar & vocabulary" have been tested through different methods, namely, "MC cloze, MC comprehension, MC paraphrase". This test consisted of 15 items for each trait and method. The overall framework of traits and methods treated is shown in the table 1 below:

**Table 1. The overall framework of traits and methods.**

Method Trait	Mc Cloze	Mc Compreh	Mc Paraph	Total
Grammar	15	15	15	45
Vocabulary	15	15	15	45
				90

Since there was no single valid test for the study. The researchers have necessarily

used Michigan and TOFEL tests as valid sources for the procedure.

### Statistical procedure

In order to investigate the research questions, a battery of statistical analyses: namely descriptive statistics, correlation analysis, multivariate analysis of variances and factor analysis have been carried

out. The major findings of the study have been presented under the above mentioned topics.

#### 1- Descriptive statistics

Table 2 has represented the mean, standard deviation, and variance for the different instruments employed in this study.

**Table 2. Descriptive Statistician**

Tests	Mean	Standard Deviation	Variance
Paraphrase Voc.	9.28	2.12	4.50
Paraphrase gram	8.27	2.44	5.95
Comprehension Voc	8.34	2.48	6.17
Comprehension gram	7.08	3.21	10.31
Cloze Voc	10.85	1.97	3.98
Cloze gram	8.28	2.56	6.56

As it has been indicated in Table 2, the highest mean observed, 10.85, belongs to the cloze vocabulary. After cloze vocabulary the highest mean score belongs to paraphrase vocabulary with 9.28 and at the third place, there comes comprehension vocabulary with the mean score of 8.34. The lowest mean observed, 7.08, belongs to the Comprehension grammar tests respectively. Cloze grammar and paraphrase grammar have nearly the same mean scores.

The comprehension grammar tests enjoy the highest standard deviation i.e.3.21 and cloze vocabulary tests enjoy lowest standard deviations

with 1.97 respectively. The same fact is observed about the variance with 10.31 for comprehension grammar tests and 3.98 for cloze vocabulary tests.

This table has revealed that there has been a meaningful relationship between the results observed for mean, standard deviation and variance in each group of tests.

#### 2- Correlation analysis;

The correlations existing between the tests have been measured through the Pearson product correlation coefficients. Table 3 displays the correlation coefficients between the tests.

**Table 3. Correlation Analysis**

	paraph. Voc	Paraph. Gram	comp voc	comp gram	cloze voc	cloze gram
paraphrase voc.	1					
paraphrase gram.	0.14	1				
Comprehension voc	0.40	0.26	1			
Comprehension gram	0.13	0.34	0.36	1		
cloze voc	0.10	0.32	0.55	0.53	1	
cloze gram	0.25	0.34	0.52	0.24	0.36	1
P < .52	Significant at .05 level of significance					

As table 3 has represented, eight correlation coefficients, out of fifteen calculated coefficients, are significant at .05 level of significance. The highest observed correlation coefficient, 0.55, exists between the cloze vocabulary and the comprehension vocabulary tests. The second high correlation coefficients, i.e. 0.53, exists between cloze

vocabulary and comprehension grammar tests and at the third place there is correlation coefficients between cloze grammar and comprehension vocabulary tests with 0.52. The fourth correlation coefficients exists between comprehension vocabulary and paraphrase vocabulary tests with 0.40 level of significance. The correlation coefficients

between cloze grammar and cloze vocabulary tests has been the same as comprehension grammar and comprehension vocabulary tests with 0.36. This fact is seen about the correlation between comprehension grammar and paraphrase grammar and cloze grammar and paraphrase grammar tests with 0.34 level of significance. For the low correlation coefficients, the lowest coefficients observed, 0.10 is that of the cloze vocabulary and the paraphrase vocabulary tests. Here is an interesting point which indicates a close correlation coefficients between paraphrase vocabulary tests with paraphrase grammar and comprehension grammar tests.

### 3- Analysis of Variance

An analysis of variance (ANOVA) has been employed to investigate the possible difference and

interaction between the variables. There has been six tests in this study that could have been grouped under two within subject factors, namely method with three levels, paraphrase, comprehension and cloze, and trait with two levels, vocabulary and grammar.

Prior to the analysis of variance, the Mauchly Sphericity test was employed to determine the homogeneity of the variance of the means. The amount of statistic calculated, .95, with a probability level of .42, has revealed that the mean enjoys enough homogeneity to be analyzed through the ANOVA.

Table 4 shows the F ratio calculated for the method, skill and method by skill factors.

**Table 4 .ANOVA Table**

Source of variations	SS	DF	MS	F	P
Within cells	273.48	68	4.02	15.48*	.000
Method	124.52	2	62.26		
Within cells	147.78	34	4.32	27.70*	.000
Skill	120.39	1	120.39		
Within cells	299.42	68	4.40	3.62*	.03
Method by skill	31.91	2	15.96		

P ≥ .05 \* Significant at .05 level of significance

The significant F ratios has indicated that the method, skill and the interaction between them have had a high impact on the means of the different tests.

Following the ANOVA, a post hoc Comparison of means through the Scheffe test was used. Table 5 has represented the results of Scheffe tests.

**Table 5. Scheffee Tests**

Comparison	X1	X2	Tobs	Tcrit	Signific
paraph x comp	7.71	8.92	3.57	2	*
paraph Voc x Par gram	9.28	8.57	1.41	2.02	
paraph voc x Voc	9.28	8.34	1.87	"	
paraph voc x gram	9.28	7.08	4.38	"	*
paraph voc x cloz voc	9.28	10.85	3.13	"	*
paraph voc x cloz gram	9.28	8.28	1.99	"	
Par gram x com voc	8.57	8.34	.45	"	
Par gram x com gram	8.57	7.08	2.97	"	*
Par gram x cloz voc	8.57	10.85	4.54	"	*
Par gram x cloz gram	8.57	8.28	.57	"	
Com voc x com gram	8.34	7.08	2.51	"	*
Com voc x cloz voc	8.34	10.85	5	"	*
Com voc x cloz gram	8.34	8.28	.12	"	
Com gram cloz voc	7.08	10.85	7.51	"	*
Com gram cloz gram	7.08	8.28	2.39	"	*
Cloz voc x cloz gram	10.85	8.28	5.12	"	*
Voc x gram	9.49	7.97	5.28	1.96	*
Voc x cloz	9.49	9.56	.34	1.96	
Gram x cloze	7.97	9.56	5.52	1.96	*

As table 5 has shown, twelve comparisons out of the nineteen made comparisons display significant difference between the means.

#### 4- Factor Analysis

To probe the underlying constructs of the implemented tests a factor analysis through varimax rotation has been run. As shown in table 6, the cloze vocabulary, comprehension grammar, and the

paraphrase grammar tests load on the first factor. Their loadings have been .81, .79 and .61 respectively. The paraphrase vocabulary, comprehension vocabulary and cloze grammar tests have loaded on the second factor; with loadings equal to .86, .68 and .59. The correlation between the two factors has been .59.

**Table 6. Factor Analysis through Varimax Rotation.**

	Factor 1	Factor 2
cloze voc	.81	
Comp gram	.79	
paraph gram	.61	
Paraph voc		.86
Comp voc		.68
Cloze gram		.59
R factor 1,2 = .59		

#### Data Analysis for question I

The correlation coefficients between the three tests of vocabulary have been as follows:

Cloze with paraphrase. 10

Cloze with comprehension. 55\*

Comprehension with paraphrase.40\*

The t- transformation of the above coefficients have been; .57, 3.8 and 2.5. Since the critical value of t at 33 degrees of freedom has been 2.02, it could be claimed that the last two correlation coefficients have statistically been significant while the correlation coefficients between the cloze and comprehension, and comprehension and paraphrase tests on vocabulary has not statistically been meaningful.

#### Data Analysis for question II

In order to probe the second question, the correlation coefficients between the three tests on grammar have been calculated as follows:

Cloze with paraphrase .34\*

Cloze with comprehension .24\*

Comprehension with paraphrase .34\*

The t- equivalent of the above coefficients have been 2.08, 1.42 and 2.08 when compared with the critical value of t at 35 degrees of freedom, it could be concluded that the first and last coefficients have statistically been significant.

Two out of the three calculated coefficients have been meaningful, while the correlation between the cloze and comprehension tests hasn't been significant.

#### Data Analysis for question III

The correlation coefficients between the vocabulary and grammar test (skills or traits), through the three methods have been:

Vocabulary paraphrase with grammar paraphrase .14

Vocabulary comprehension with grammar comprehension .36\*

Vocabulary cloze with grammar cloze, 36.

The t- transformation of these coefficients have been; .81, 2.21 and 2.21. The critical value of t at 33 degrees of freedom has been 2.02. Thus it can be concluded that the correlation between the paraphrase vocabulary and grammar tests was the only non-significant coefficient.

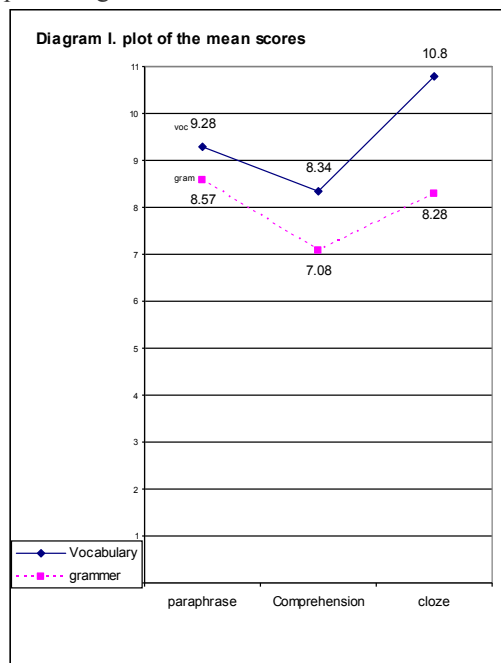
#### Data Analysis for question IV

To probe the possible differences between the tests, an analysis of variance (ANOVA) has been carried out. The F ratios for the method, trait and the interaction between them (Method by trait) have been 15.48, 27.70 and 3.62 which were all significant at .05 level of significance. In other words, the method, trait and the interaction between them has had a significant impact on the student's performance on the tests.

A close investigation of the results has revealed the fact that the overall mean score of the vocabulary test has been higher than the mean score for grammar. In both traits, the mean scores of the paraphrase method have been higher than those of the comprehension method, after which the mean scores increase for the cloze method.

A factor analysis through varimax rotation has been employed to answer the fourth question. The results have shown that the cloze vocabulary, comprehension grammar and paraphrase grammar tests have loaded on the first, and the paraphrase vocabulary, comprehension vocabulary and cloze grammar tests have loaded on the second factor.

Diagram 1 has represented the pattern of the means plotted against the method and trait factors.



### Conclusion

These findings have been consistent with findings of Handel's (1987) study which suggested that two commonly used measures of college outcome yield scores that are reasonably high indices of convergent validity. The highest correlation observed between the three tests of college outcome. The findings of the study have shown that distinct traits within a given method are moderately inter-correlated.

Correlations of the three factors as scored by different scorers have been studied by a multitrait-multimethod procedure. The results provided strong support for both convergent and discriminant validity of three factors. Having in mind the mentioned questions and problems the researchers used a multitrait – multimethod matrix.

A hetro-method – monotrait analysis has been conducted to find answers for the first two developed questions. The findings has shown that there has been a strong correlation among methods applied. A heterotrait- heteromethod analysis has been used to answer the third developed question, and the result has shown that there has been a strong correlation among methods and traits.

Finally, the researchers have found that there has been an underlying construct for traits (grammar and vocabulary) and methods (MC cloze, MC paraphrase, and MC comprehension), in the conducted experiment.

### Pedagogical Implications

Language testing is central to language teaching. It provides goals for language teaching and monitors for both language teachers and learners success in reaching those goals. It has a very strong influence on methodology.

Language testing is a complicated issue and much of these complications come from problems of description and measurement which are particularly acute in linguistic and psychological investigations.

This study has clearly shown that performance on language tests is influenced by at least two independent factors: the effect of test method and the effect of the traits being measured. As it was shown in the process of study there has been general agreement among the six tests in ranking the subjects across traits and methods.

Since language teaching and language testing have backwash and wash back effect on each other, it is due to the teachers and test practitioners to make use of different methods in their testing methodologies. It is also helpful to the material developers to benefit from the findings of this study when designing materials for teaching and testing.

### Suggestions for Further Research

The research discussed here only has begun to scratch the surface. Many unanswered questions remain, and many of the answers proposed, will no doubt need to be replicated and refined if not in fact discarded.

Given the importance of testing in language teaching and the significant role of the testing methods in reading comprehension, strategies of testing methods used for this purpose will be a fruitful area for further research.

The sample population chosen as subjects for this study was only male students. One can broaden the scope of research to a large extend to include female participants in different universities.

The same investigation may give more insights to language testing instruction if it is carried out with the lower intermediate and advanced level of language learners.

The same procedure can be applied for more traits and methods as well as different language skills like listening comprehension. It can also be applied to other fields of science like physics and chemistry.

### Corresponding Author:

Sohrab Rezaei  
 Assistand Professor Allameh Tabatabaei University  
 Tel:+982122195144  
 Mobile:+989363224553  
 E-mail:[Sohrab\\_rezaei@yahoo.com](mailto:Sohrab_rezaei@yahoo.com)



**BIBLIOGRAPHY**

1. Baker David (1989) language testing. A critical survey and practical Guide. University of Bahrain.
2. Bennett Tepper et al (1993) Discriminant and Convergent Validity of the problem Solving Style Questionnaire Educational and Psychological Measurement.V 53,PP 434-444
3. Bresna Han, Mary J, et al.(2005), A Multitrait-Multimethod validity assessment of self construal in Japan, Korea and the U.S. Human Communication Research; V 31, N 1, PP 33-59.
4. Campbell, D.T, and Fiske, D.W. (1959) convergent and discriminant validation by the multitrait – multimethod matrix. Psychological Bulletin.
5. Campbell, D.T, and O'Conneil, E, J. (1967). Method factors in multitrait – multimethod matrices. Multiplicative rather than additive. Multivariate Behavioral Research. V2, PP 409-426.
6. Chang Lie, (1993), Using confirmatory factor analysis of MT-MM data to assess the psychometrical equivalence of 4-point and 6-point Likert-type scales.
7. Darwin D.Hendel (1991), Evidence of Convergent and Discriminant Validity in three Measures of College Outcome. PP 351-357.
8. Davies Alan (1990) Principles of language testing. Basil Blackwell publisher, Cambridge, USA.
9. Douglas N. Jackson. (1975) Multimethod Factor Analysis: A Reformulation Multivariate Behavioral Research. PP 259– 276.
10. Farhady. H, Jafarpoor, A, and Birjandi P. (1994) Language skills Testing from theory to practice, SAMT.
11. Farr, Roger and Jongsma, Eugene (1993) the Convergent, Discriminant Validity of integrated reading, writing assessment. Journal of Research and Development in Education.
12. Henning Grant (1987) A Guide to Language Testing Development Evaluation Research, Newbury House publishers University of California, Los Angles.
13. Harris P.David (1990), Testing English as a Second language MC Graw Hill book Company.
14. Herbert W.Marsh & Dennis Hocevar, (1983) Confirmatory Factor Analysis of Multitrait-Multimethod Matrices.20, 231-248.
15. -Joseph Levin (1989). Multiple Group Factor Analysis of Multitrait - Multimethod Matrices. Multivariate Behavioral, Research, 23,469-487.
16. Joseph Stevens and Patricia, S .Clauser, (2005). Multitrait-Multimethod comparison of selected and constructed response, Assessment of language achievement. University of New Mexico.
17. Kavanaough, M.J.A.C Mackinney and L Wolins (1991) Issues in managerial performance: Multitrait – Multimethod analysis of Ratings. Psychological Bulletin 75, 34-49.
18. Kellabery, A.L. and R. Kluegel (1975) Analysis of the Multigrain.
19. Multimethod Matrix. Some Limitations and an Alternative. Journal of Applied Psychology 60:1-9.
20. Lorena Loisa, (2007). Validating a standard-based classroom assessment of English proficiency. Journal of Language Testing, V, 24, N. 4, PP 489-515.
21. Lowe – NK Ryanwenger – NM, (1992) Beyond Campbell and Fiske- Assessment of convergent and Discriminant validity, Research in Nursing & Health, 15, 67-75.
22. Lomax, R.G. and J.Algina (1979). Comparison of two procedures for analyzing Multitrait- Multimethod matrices. Journal of Educational Measurements 16,177-186.
23. Madsen S. Harolds, (1983) Techniques in testing. Oxford, England.
24. Muis Krista R. (2007). Using MT-MM analysis to examine conceptual similarities of three self regulated learning inventories. British Journal of Educational Psychology. V. 77, PP.177-195.
25. Oller, John W.Jr, (1979) language tests at school. A pragmatic Approach. Longman group limited, London.
26. Plamer, s Adrian and Groot, M.J. peter. The construct validation of test of communicative competence.
27. Pike G. and Banta T. (1989), Using construct validity to Evaluate Assessment Instruments. Research Report.
28. Robert J. Vandenberg & Vida Scarpello (1991) Multitrait- Multimethod Validation of the satisfaction with my supervisor scale; Educational and Psychological Measurement 51,203-212.
29. Ray, M.L, Heeler, R.M. (1975) Analysis techniques for exploratory use the multitrait multimethod matrix Educational and Psychological Measurement, 35,255-265.
30. Roger Farr Evgene jonsma (1993), the convergent, Discriminant validity of Integrated Reading/Writing assessment. Journal of Research and Development in Education. 26,83-90.
31. Schmitt, N, Coyle, B.W. & Saari, B.B. (1997) A review and critique of analyses of multitrait – multimethod matrices. Multivariate Behavioral Research, 12,447-478.
32. Schmitt, N, et al, (1977), A review and critique of analysis of multitrait – multimethod Matrices. Multivariate Behavioral Research, 12:442-778.
33. Stephen L. Goling & Edward Seidman (1974) Analysis of Multitrait- Multimethod Matrices: A two step principle components procedure. Multivariate Behavioral Research, 9,479-496.
34. Susanna Martha Brutsh (1980) convergent, discriminant validation of proficiency in oral and written production of French. Paper presented at the TEOSL.
35. Vinary Kanetkar, Martin G. Evans (2001). Confirmatory factor analysis of MT-MM matrices. University of Toronto.
36. [WWW.Socialresearchmethods.net](http://WWW.Socialresearchmethods.net).
37. [WWW.sagepub.com](http://WWW.sagepub.com).
38. [WWW.uoregon.edu](http://WWW.uoregon.edu).
39. [WWW.ingentaconnect.com](http://WWW.ingentaconnect.com).
40. [WWW.wikipedia.org](http://WWW.wikipedia.org).