Vollmer, H.J. (1983). The structure of foreign language competence. In A. Hughes & D. Porter (Eds.), Current developments in language testing (pp. 3-30). UK.: University of Reading.

Weir C.J. (1990). Communicative language testing. C.N. Candling (Ed.), Language teaching methodology series (pp. 1-21). Great Britain: Prentice Hall.

Widdowson, H.G. (1971). The teaching of students of science and technology. In Science and technology in a second language. Center for Information on Language Teaching.

Oller, J.W. (Ed). (1983 d). Issues in language testing research. Rowley, Mass.: Newbury House.

Oller, J.W. & Hinofotis, F.B. (1976). Two mutually exclusive hypotheses about second language ability: Factor analytic studies of a variety of language tests. Paper presented at the annual meeting of the Linguistic Society of America, December, 1976.

Oller; J.W. & Hinofotis, F.B. (1980). Two mutually exclusive hypotheses about second language ability: indivisible or partially divisible competence. In J.W. Oller & K. Perkins K. Perking (Eds.) Research in language testing (pp. 13-23). Rowley, Mass.:Newbury House.

Palmer, A.S. & Groot P.J.M. (1981). Introduction to test valibity. In A.S. Palmer; P.J.M. Groot & G.A. Trosper (Eds.) The construct validation oftests of communicative competence. Washington, D.C.; TESOL.

Song, F. et al. (1986). Models of second language competence: A structural equation approach. In Language testing, Vol. 3, No. 1, 54-79.

Savignon, S.J. (1972). Communicative competence: An experiment in foreign language teaching. Philadelphia Pa: Center for Curriculum Development.

Spolsky, B. (1973). What does it mean to know a language, or how do you get someone to perform his competence? In J.W. Oller & J.C. Richards (Eds.), Focus on the lesrner: Pragmatic perspectives for the langusge teacher (P. 174). Rowley, Mass.: Newbury House.

Spolsky, B. (Ed.). (1978). Approaches in language testing. Series 2. In Arlington Va. Center for Applied Linguistics.

Spolsky, B. & Jones, R.L. (1975). Testing language proficiency (p. 69). In Arlington Va: Center for Applied Linguistics.

Harris, D.P. (1969). Testing English as a second language (pp. 5, 34). New York: McGraw-Hill

Hughes, A. & Porter, D. (1983). Current developments in language testing. (pp. 3-37). UK. University of Reading.

Hindmarsh, R. (1980). Cambridge english lexicon. combridge University press.

Howell, D.C. (1989). Fundamental statistics for the behavioral science. Boston: PWS-KENT, Publishing Co.

Kelly, R. (1978). On the construct validation of comprehension tests: an exercise in applied linguistics. Ph.D. thesis, University of Queensland.

Kunnan, A. (1999) Validation in language assessment. OUP.

Lado, R. (1961). Language testing: The construction and use of foreign language tests (p.2). New York: McGraw-Hill Book Co.

Milanovic, M. (1988). The construct validation of a performance based battery of English language progress test. Ph.D. thesis, University of London.

Morrow, K.E. (1979). Communicative language testing: revolution or evolution? In C.J. Brumfit & K. Johnson (Eds), The communicative approach to language teaching (p.145). London: OUP.

Oller, J.W. (1977). The psychology of language and contrastive linguistics: the research and the debate (p. 13). Foreign Language Annals.

Oller, J.W. (1978). The language factor in the evaluation of bilingual education. In J. Alatis (Ed.), International dimension of bilingual education. Washington, D.C.: Georgetown University Press.

Oller, J.W. (1976). language tests at school (pp. 212, 425). London: Longman.

Oller, J.W. (1980). A comment on specific variance versus global variance in certain EEL tests. TESOL Quarterly, 14(4), 527-30.

Testing and Research, Examining Modern Language" (pp. 38-45). London: CILT & Committee on Research and Development in Modern Languages (CILT Reports & Papers,4).

Cummins, J. (1983). Language proficiency and academic achievement. In J.W. Oller (Ed.), "*Issues in language testing research*" (pp. 108-127). USA: Newbury House.

Eble, R.L. (1972). "*Essentials of educational measurement*". Englewood Cliffs, NJ: Prentice-Hall, Inc.

Farhady, H. (1980) "*Justification, development, and validation of functional language tests*". Unpublished doctoral dissertation, University of California, Los Angeles.

Farhady, H. (1983a). On the plausibility of the unitary language proficiency factor. In J.W. Oller (Ed.), "*Issues in language testing research*" (pp. 11-29). USA: Newbury House.

Farhady, H. (1983b). The disjunctive fallacy between discrete-point and intergrative tests. In J.W. Oller (Ed.), "*Issues in language testing research*" (pp. 311-323). USA: Newbury House.

Farhady, H. (1995). Research methods in applied linguistics. Tehran: Payame-Noor University.

Farhady, H., Jafarpoor A. & Birjand, P. (1994). Testing language skills. Tehran: SAMT.

Fries, C.C. (1945). Teaching and learning English as a foreign language. Ann Arbor: University of Michigan Press.

Harley, B. et al. (1990). The development of second language proficiency. OUP

Alderson, J.C. (1981). Report of the discussion on general language proficiency. In. J.C.

Alderlson & A. Hughes (Eds.), *"Issues in language testing"* ELT Documents 111 (p.187). London: The British Council.

Aldesron, J.C. (1986). Innovations in language testing? In M. Portal (Ed), *"Innovations in language testing"*. Windsor: NFRE-Nelson.

Alderson, J.C. (1991). Language testing in the 1990s: How far have we come? How much further have we to go? In J.C.

Anivan (Ed.), *"Current developments in language testing"* (pp. 1-19). Singapore: SEAMEO Regional Language Center.

Bachman, L.F. (1990). *"Fundamental considerations in language testing"*(pp. 81-108). OUP.

Bachman, L. and Palmer, A. (1996) Language in language tests in practice. OUP

Baker, D. (1989) *"Language testing: a critical survey and practical guide"* (p. 72). Great Britain: British Library Cataloging in Publication Data.

Canale, M. (1983). On some dimensions of language proficiency. In J.W. Oller (Ed.), *"Issues in language testing research"* (p. 339). USA: Newbury House.

Carroll, J.B. (1961a). Fundamental considerations in testing for English language proficiency of foreign students. In *"Testing the English proficiency of foreign students"* (pp. 30-40). Washington, DC: Center for Applied Linguistics.

Carroll, J.B. (1983). Psychometric theory and language testing. In J.W. Oller (Ed.), *"Issues in language testing research"* (p. 104). USA: Newbury House.

Chaplen, E.F. (1970). *"Oral examinations. In Center for Information on Language*

different levels. They may be treated atomistically at the lower levels but holistically at the higher levels, The same holds true for regard to testing language proficiency.

As a closing note, it looks reasonable to resort to Alderson's (1991) quotation when he states:

> *Language proficiency is a complex phenomenon, and is very little understood despite the best efforts of many social science disciplines to attempt to elucidate it. There are many different, and indeed sometimes competing models of language proficiency, and we are barely at the beginning of operationalizing and testing and validating those models.x Research into testing models is fairly recent, and has a long way to go yet: there is a lot we do not know, a frightening amount of ground to be covered... there centainly is a lot more to be made.*

Then, is seems obvious that the first necessity will be the replication of the study. Furthermore, many other factors such as the homogeneity of subjects, amount of their schemata, test format, and the teaching method should be considered as effective variables in further studies. Last, but not least, the "Level-dependent Approach" is in its infancy which requires to be brought up through further research studies

References

Alderson. J.C. (1978). A study of the cloze procedure with native and non-native speakers of English. unpublished Ph.D. thesis, University of Edinburgh.

In sum, the data from both the correlation coefficient matrices and factor loadings indicate that language proficiency should not be perceived of as having a stable and constant underlying pattern. Rather, language ability may be a Level-dependent phenomenon in nature. It means that language proficiency can be multifactorial at the lower levels but more unifactorial at the higher levels. Finally, the results regarding the second research question raise more questions than they answer. When tests types are compared by individual correlation coefficient, differences are significant but when they are studied in terms of mean of the correlations coefficients, they look rougly interchangeable.

Implications

The findings of this study have some theoretical as well as pedagogical implications. Theoretically, the ascending order of the correlation coefficiets and reduction of the number of the traits from three at the elementary and intermediate levels to two at the advanced level leads to a new theoretical framework which can called **Level-dependent Approach.** This framework, implicitly mentioned by Vollmer, Milanovic, and Anivan (1990), may help scholars formulate a more descriptive theoretical framework for language ability. It is also plausible to assume that different theoretical frameworks should be formulated to accommodate the nature of language proficiency based on factors influencing its structure.

Pedagogically, the proposed, level-dependent theoretical framework implies that language skills and components should be treated differently at

Table 12
Varimax Rotated Factor Loadings on the Study Variables at the Advanced Level

Variables	Factors 1	2	h^2	reliability
LISCOM	0.87	*	0.82	0.89
LISCLOZ	0.70	*	0.53	0.73
GRAMWRIT	*	0.86	0.78	0.82
READCOM	0.59	0.62	0.73	0.80
READCLOZ	0.32	0.78	0.71	0.79
VOC	*	0.81	0.67	0.77
PRO	0.60	*	0.42	0.57
DICT	0.81	*	0.65	0.80
WRITCLOZ	0.74	0.36	0.68	0.82

* Loadings below 0.30

As the Tablse show, there are heavy loadings on factor 2 from listening skill and on factor 3 from grammar and vacabulary at the elementary and interrmediate levels. But except for vocabulary, the remaining variables at these two levels loaded heavilty on factor 1. Reading comprehension skill at the elementary level shows the multifact orial nature of language proficiency. The loadings were so dispersed that it was too difficult to lable the factors at the first two levels. However, the number of factors generated at the advanced level was different from those of the other two levels, since it was constrained to only two factors. Then, due to the loading of oral skills such as listening comprehension, pronunciation and dictation tests on the factor 1 at this level, it may by called oral language ability factor. While reading comprehension, vocabulary and grammar loaded more on factor 2 which may be identified as reading ability.

Table 10

Varimax Rotated Factor Loadings on the Study Measures at the Elementary Level Factors

	Factors				
Variables	1	2	3	h^2	reliability
LISCOM	*	0.79	*	0.64	0.79
LISCLOZ	*	0.55	*	0.36	0.59
GRAMWRIT	*	*	0.81	0.74	0.74
READCOM	0.38	0.71	*	0.72	0.85
READCLOZ	0.65	*	*	0.50	0.60
VOC	0.50	*	0.59	0.60	0.66
PRO	0.63	*	*	0.45	0.56
DICT	0.87	*	*	0.79	0.82

* Loadings be;pw 0.30

Table 11

Varimax Rotated Factor Loadings on the Study Measures at the Intermediate Level

	Factors				
Variables	1	2	3	h^2	reliability
LISCOM	0.40	0.65	*	0.58	0.72
LISCLOZ	0.31	0.68	*	0.56	0.72
GRAMWRIT	0.35	*	0.72	0.66	0.70
READCOM	0.85	*	*	0.80	0.90
READCLOZ	0.81	*	*	0.76	0.86
VOC	*	*	0.81	0.67	0.73
PRO	0.71	*	*	0.64	0.76
DICT	0.77	*	*	0.63	0.77
WRITCLOZ	0.54	*	0.63	0.70	0.74

* Loadings below 0.30

Table 9

Cross-Comparison of the Mean of Correlation Coefficients among the Test Types at the Three Proficiency Levels

Level	Test Type	Mean
Elementary	DP	0.39
	IN	0.37
	SemiDP/IN	27
Intermediate	DP	0.44
	IN	0.43
	SemiDP/IN	0.27
Advanced	DP	0.44
	IN	0.50
	SemiDP/IN	0.34

Although correlation coefficients provide useful information on the go to getherness of the scores at different levels of language proficiency, they do not indicate the underlying sameness of the abilities being measured (Farhady, 1980, 1983). To this end, several factor analyses were performed to indentify the underlying factors present in each of the proficiency levels. The factor analyses were conducted with varimax rotation to avoid any methodological problems (Farhady, 1983; Comrey, 1975). The results of factor analyses along with the reliability coefficients are reported in Table 10, 11, and 12 for different levels of language proficiency.

Table 8

Cross Camparison of Correlation Coefficients at the Three Proficiency Levels

Level Variavles		Elementary	Intermediate	Advanced
Liscom	Liscloz	0.43	0.58	0.65
Liscom	Gram	0.25	0.35	0.38
Liscom	Readcom	0.53	0.53	0.68
Liscom	Readcloz	0.20	0.46	0.47
Liscom	Voc	0.10	0.24	0.30
Liscom	Writcloz	-	0.42	0.73
Liscom	Pro	0.08	0.54	0.58
	Dict	0.17	0.62	0.72
Liscloz	Gram	0.25	0.26	0.31
Liscloz	Readcom	*0.50	0.42	0.54
Liscloz	Readcloz	0.25	0.37	0.39
Liscloz	Voc	*0.18	0.15	0.23
Liscloz	Writcloz	-	0.30	0.59
Liscloz	Pro	0.15	0.52	0.47
Liscloz	Dict	0.25	0.58	0.58
Gram	Readcom	0.44	0.48	0.65
Gram	Readcloz	0.24	0.52	0.73
Gram	Voc	0.51	0.64	0.72
Gram	Writcloz	-	0.65	0.46
Gram	Pro	0.22	0.35	0.33
Gram	Dict	0.20	0.33	0.20
Readcom	Readcloz	0.43	0.79	0.67
Readcom	Voc	*0.36	0.29	0.56
Readcom	Writcloz	-	0.61	0.66
Readcom	Pro	*0.31	0.24	0.50
Readcom	Dict	*0.46	0.37	0.51
Readcom	Voc	*0.46	0.37	0.66
Readcom	Writcloz	-	0.65	0.52
Readcloz	Pro	*0.46	0.22	0.38
Readcloz	Dict	*0.61	0.33	0.30
Readcloz	Voc	0.46	0.37	0.66
Readcloz	Writcloz	-	0.65	0.52
Readcloz	Pro	*0.46	0.22	0.38
Readcloz	Dict	*0.61	0.33	0.30
Voc	Writcloz	-	0.59	0.37
Voc	Pro	*0.44	0.34	0.26
Voc	Dict	*0.51	0.25	0.26
Writcloz	Pro	-	0.32	0.53
Writcloz	Dict	-	0.33	0.62
Pro	Dict	0.35	0.62	0.50

* = Unexpected correlation coefficients

As the correlation matrices and particularly Table 8 indicate, there is an ascending trend in the degree of go-togetherness of the scores from the elementary to the advanced levels. Then, except for a few number of measures, the advanced level showed the highest correlation coefficients and the two other levels stood in the lower positions, respectively.

Out of the thirty-six correlations at the intermediate and advanced levels, except for a few variables, the advanced level enjoyed 25 higher correlations than those of the intermadiate level and 24 higher correlation than those of the elementry level. Therefore, one could argue that there is relationship between the discreteness and the tntegrativeness of language ability not in terms of the language abilities themselves but in terms of the levels of language ability. That is, at the lower levels of language proficiency, language components seem to be discrete and separete from one another. As the language proficiency increases, however, so does the interagtion of the language components into a holistic ability. This relationship can be holistically observed by examining the cross-comparisions of the mean correlation coefficients among the test types at each level presented in Table 9. These finding are in line with those reported in literature that the structure of language ability depends, to some extent, on the degree of language proficiency.

Table 5
Correlation Coefficients among the Study Measures at the Elementary Level

Subtests	1	2	3	4	5	6	7	8
LISCOM (1)	*							
LISCLOZ (2)	0.43	*						
GRAMWRIT (3)	0.25	0.25	*					
READCOM (4)	0.53	0.50	0.44	*				
READCLOZ (5)	0.20	0.25	0.24	0.43	*			
VOC (6)	0.10	0.18	0.51	0.36	0.46	*		
PRO (6)	0.08	0.15	0.22	0.31	0.46	0.44	*	
DICT (8)	0.17	0.25	0.20	0.46	0.61	0.51	0.35	*

Table 6
Correlation Coefficients among the Study Measures at the Intermediate Level

Subtests	1	2	3	4	5	6	7	8	9
LISCOM (1)	*								
LISCLOZ (2)	0.58	*							
GRAMWRIT (3)	0.35	0.26	*						
READCOM (4)	0.53	0.42	0.48	*					
READCLOZ (5)	0.46	0.37	0.52	0.79	*				
VOC (6)	0.34	0.15	0.64	0.29	0.37	*			
PRO (7)	0.54	0.52	0.35	0.24	0.22	0.34	*		
DICT (8)	0.62	0.59	0.33	0.37	0.33	0.25	0.62	*	
WRITCLOZ (9)	0.42	0.30	0.65	0.61	0.65	0.59	0.61	0.33	*

Table 7
Correlation Coefficients among the study Measures at the Advanced Level

Subtests	1	2	3	4	5	6	7	8	9
LISCOM (1)	*								
LISCLOZ (2)	0.65	*							
GRAMWRIT (3)	0.38	0.31	*						
READCOM (4)	0.68	0.54	0.65	*					
READCLOZ (5)	0.47	0.39	0.73	0.67	*				
VOC (6)	0.30	0.23	0.72	0.56	0.66	*			
PRO (7)	0.58	0.47	0.33	0.50	0.38	0.26	*		
DICT (8)	0.72	0.58	0.20	0.51	0.30	0.26	0.50	*	
WRITCLOZ (9)	0.73	0.59	0.46	0.66	0.52	0.37	0.53	*	

Table 3
Descriptive Statistics for the Study Measures at the Intermediate Level

Subtests	X	S
LISCOM	12.09	3.29
LISCLOZ	9.55	2.56
GRAMWRIT	13.86	4.08
READCOM	14.02	3.37
READCLOZ	21.08	6.12
VOC	14.55	3.36
PRO	15.57	3.54
DICT	60.08	6.14
WRITCLOZ	13.55	3.32

WRITCLOZ = Writing Cloze Task

Table 4
Descriptive Statistics for the Study Measures at the Advanced Level

Subtests	X	S
LICOM	14.23	3.78
LISCLOZ	12.50	3.5
GRAMWRIT	27.30	6.69
READCOM	13.06	4.32
READCLOZ	13.74	5.49
VOC	12.77	4.14
PRO	13.63	3.19
DICT	94.91	7.92
WRITCLOZ	16.26	2.62

To determine the intercorrelations of the skills and components of language proficiency and to investigate the similarities between DP, IN, and Semi DP/IN type tests at the three proficiency levels, Pearson Product Moment Correlation was utilized. Tables 5, 6, and 7 present the obtained results.

of testing. Furthermore, the analysis were done on each level of language proficiency separately in order to investigate the differences at the three levels.

Results

Tables 2, 3 and 4 present the descriptive statistics of the subtests at the intended levels.

Table 2

Descriptive Statistics for the Study Measures at the Elementary Level

Subtests	X	S
LISCOM	8.21	2.26
LISCLOZ	5.69	2.56
GRAMWRIT	16.50	3.15
READCOM	15.90	4.13
READCLOZ	10.01	3.40
VOC	14.70	3.24
PRO	12.67	3.14
DICT	44.99	5.55

LISCAM=Listening Comprehension, LISCLOZ=Listening Cloze,
GRAMWRIT=Grammar & Written Expressions,
READCOM=Reading Comprehension, READCLOZ=Reading Cloze,
VOC=Vocabulary, PRO=Pronunciation, DICT=Dictation

Procedure

In order to make data collection procedure appropriate to the variables under investigation, certain steps were followed. First, TOEFL was administered to about 500 students. Based on their scores, and following the principles of normal distribution, the students whose scores ranged between plus or minus one standard deviation from the mean were taken as the intermediate group; the students whose score were one standard deviation above the mean were taken as advanced group; the students whose scores were one stsndard deviation below the mean were taken as the elementary group; and the students whose scores were beyond 95% probability on either side of the normal distribution were discarded as outliers. Eliminating the outliers which amounted to 77 subjects, the remaining 423 subjects performance was included in data analysis.

As for scoring, each correct answer was given one point. Dictation type tests were scored using the word-for-word method in which every morphologically independent element was considered an item and given credit, and the extra words inserted in the text were taken as wrong. Cloze type tests were scored utilizing the exact word method which is claimed to be appropriate for foreign language context. (Farhady et al., 1994).

Analysis

Data analysis included descriptive, Correlational, and factor analyses. To conduct correlational and factor analysis, the test scores were classified into three categories of DP, semi-IN and IN to first follow the traditional DP-IN continuum, and second, to have a framework for skills and methods

fourth, the materials for reading comprehension tests were selected using Fog's index of readability applied to the materials concerned with students educational courses. Although some reservations exist on the use of readability scales, Fog's index seemed practical for this study. Table 1 presents the number of items, the test methods, and the readability levels (wherever applicable) of the tests used for each level.

Table 1

Specifications of Test Battery

Level	Elementary		Intermediate		Advanced	
Subtests	Number of Items	Readability Level	Number of Items	Readability Level	Number of Items	Readability Level
Listening Comprehension	10	--	17	-	17	-
Listening Cloze	12	9.08	12	16.79	18	19.58
Reading Comprehension	20	11.86 (A) 11.54 (B)	16	15.54 (A) 15.92 (B) 16.35 (C)	20	21.96 (A) 21.88 (B) 22.5 (C)
Reading Cloze	20	13.25	30	17.44	25	24.55
Writing Cloze Task	-	-	18	-	18	-
Grammar & Written Expressions Vocabulary	20 20	- -	20 19	- -	40 20	- -
Pronunciation	20	-	20	-	20	-
Dictation	52#	10.4	70#	17	109#	19
TOTAL	174	12	222	16	287	22

* Average readability level of the materials in the curriculum

\# Each word was considerend as an item in terms of scoring

methods, and varieties of tests as consistent across the groups as possible. The only difference which was manipulated was the complexity of the language of the tests to suit the levels of proficiency.

A few points need to be mentioned here to clarify the instrumentation procedures of the study. First, an attempt was made to utilize different methods of tests. The tests were selected to represent the extreme positions on the so-called discrete-point/integrative/pragmatic continuum, i.e., from sound discrimination to cloze type test. For example, listening comprehension items ranged from the items measuring students' ability in comprehending sentences to comprehending full passages. In addition, a dictation test, to represent a different method of measuring listening comprehension ability, was administered to every group. For reading comprehension tests, in addition to the items similar to those used for listening comprehension, a cloze test was administered to represent a different method of testing. And finally, tests of language components included items ranging from measuring students sound discrimination ability to their command of the structure of the English language.

Second, most of the tests were selected from among the already validated or standardized tests. For example, for the reading comprehension part at the sentence level some of the items developed by Yorio and Oller (1983) were used. Third, the selection of the vocabulary range was based on the frequency list provided by Hindmarsh's Cambridge Lexican (1980). This source was used because it lists 4500 most frequent words in English which are classified at different levels of proficiency. And

language ability can be abtained which would help theoreticians to formulate more reasonable theories. It is hoped that this study will be a stop in the right direction. To shed some light on the issue, this research was designed to answer the following questions:

1. Is there any relationship between the levels of language ability and the degree of the integration of language skills and language components?

2. Is there any relationship between discrete point and integrative testing approaches on the one hand, and the levels of language ability on the other?

Method

Subjects

The subjects were about 500 randomly selected undergraduate students of the English language at different universities in Iran. Based on their performance on a 1995 original version of TOEFL, they were divided into three groups of elementry, intermadiate, and advanced level language proficiency groups, with 141 subjects for each group.

Instrumentation

Two instruments were used in this study. The first one was an original version of TOEFL employed as the critetion measure for determining the proficiency levels of the subjects. Second, batteries comprising tests of different language components and skills adjusted for the level of proficiency were prepared. An attempt was made to keep the forms,

performance of the language testees, which in turn, would influence the outcome of the theoretical explanations. For instance, it would be reasonable to ask questions regarding the underlying structure of language proficiency. The first set of questions deal with whether the form of the items such as true-false, multiple choice, fill in, or matching would lead to the same underlying structure of the performance. Another set of questions focused on whether the nature of the language element being maesured such as discrete-point, integrative, or pragmatic-communicative would provide the same underlying structure. The third set of questions centered on whether social, political, educational factors would influence the test takers performance and thus the nature of the construct. The last set of questions dealt with whether the level of the language ability of the learners, i.e., elementary, intermediate, and advanced, would result in similar underlying pattern. Most important of all, researchers have been interested in whether there is any interaction among these factors regarding the patterning of the language ability.

Although some research has been done to answer some aspects of the question, a systematic approach cannot be substantiated in the literature. The major outcome of research along this line would have a significant contribution to our understanding of the structure of language ability that would, in turn, help improve our models. In other words, by accounting for factors such as test method, language abilitity level, and others, we might be able to refine the language ability models from nonlanguage factors. By accounting for the contribution of nonlanguage factors, the purer nature of

Theoretically, the hypothesis was challenged by Alderson (1981) when he claimed that accepting one underlying proficiency factor would lead to the assumption of no difference among different knowledge components. But observing some other factors besides the g factor resulted in the modification of Oller's initial position. Thus, two moderate versions analogous to those of the multidimensional were speculated. Its strong version was based on the existence of only one general language proficiency factor that explains all common variance, and its moderate version was based on general factor that is common to all language abilities, and one other factor specific to the ability being measured. In spite of the modifications in the UCH, and even in the post UCH era, still a less clear picture of language proficiency remained (Alderson, 1991). These studies provide support for idea that different statistical methods are useful and, at the same time, important in investigating the nature of language proficiency (Harley, et al, 1990).

In the late 80's and early 90's a shift was made in language testing to focus on performance testing. Performance testing has trapped within the labyrinth of modeling. Reviewing several models of performance testing, McNamara cites several problems for such models. No matter what position is taken, i.e, whether competence or performance models, practitioners still face the initially formulated question by Spolsky some 30 years ago that "What does it mean to know a language?" or the queation asked by Briér "Are we really measuring language proficiency?"

One useful strategy has been to investigate the factors that influence the

The multidimensional model had two versions. The strong version which assumed 16 components for the total language proficiency, and the weak version which speculated four skills as the dimensions of language abilety (Vollmer, 1983). This multidimensional model was criticized on the grounds that it failed to accommodate for the relationship among the components and skills (Bachman, 1990). It also ignored the full context of discourse and the situation of language use (Vollmer and Sang, 1983).

In reaction to the multidimensional model, and to alleviate some of its deficiencies, Spolsky (1973) offered the concept of the overall language proficiency which motivated Oller to suggest the notion of Unitary Competence Hypothesis (UCH). According to the UCH, one underlying ability, referred to as the g factor accounts for language proficiency.

The UCH was criticized for both methodological and teoretical drawbacks. Methodologically, two major approaches have been dominantly used: exploratory and confirmatory factor analytic methods. Although each has its own variations, generally speaking, the former assumes no theoretically predetermined framework and lets the analysis identify the best fitting pattern for the construct. The latter, however, assumes a theoretically defined model and tries to test it through the statistical analysis. For instance, Oller's advocacy of using Principal Component Analysis (PCA) rather than Principal Factor Analysis (PCA) was questioned by Farhady (1983), Vollmer and Sang (1983) in that this technique would allow the incorporation of error variance into the analysis and then the overestimation of the first factor.

underlies internal consistency coefficients.

And finally, psychometric-cummunicative period was the result of dissatisfaction with the previous attempts. One of the outstanding features of this approach was the strong belief thet testing language, or any ability for that matter, would not be possible without formulating a theoretical framework for that ability. In other words, in oder to measure communicative ability, a theoretical formulation of communicative competence should be development first. Therefore, various models of communicative competence were offered. (Canal and Swain, 1980; Canal, 1983; Bachman, 1990). Thus the communicative language testing (CLT) which seemed to utilize authentic conditions of language use and capacity of the learners (Widdowson, 1983) along with contextual features and culltural considerations about language received a great amount of attention (Lynch, 1996).

Along with theoretical developments, attempts have been made to provide operational definitions of language proficiency, communicative competence and their components. Although theoretical and operational definitions of language proficiency have contributed to our understanding of the concept, and the models have provided potentially useful framework for the design of language tests (Weir, 1990), there has not been an agreed-upon definition as yet. Some of these definitions have led to the development of language ability models such as multidimensional and unidimensional ones. The outcome of the former was the so-called discrete point tests and that of the latter was the integrative and pragmatic tests.

test methods were translation, essay writing, and grammatical description. Psychometic-structuralist priod coincided with the involvement of linguistics and psychology in language education which resulted in the measurement of language skills and components through the development of discrete-point type tests (Lado, 1961; 1988). Accordingly, major test types were those of spelling, auditory discrimination, vocabulary, and grammetical structures. Contrary to their objectivity of scoring (Harris, 1969; Heaton, 1989; Baker, 1989; Weir, 1990), they suffered from several drawbacks. They were criticized on their reliability and validity (Carroll, 1961, 1978; Farhady, 1983; Weir, 1990), on their denaturalization of language (Chaplen, 1970; Savingnon, 1972), and on their ignoring of the extra linguistic factors.

Integrative-sociolinguistic period was the offspring of a triple interplay of linguistic, psychological and sociological views. The theory evolving out of these three dimensions assumed that knowledge of a language is more than just the sum of a set of discrete elements (Carroll, 1961; Spolsky, 1978). This approach, called integrative (IN) testing, tried to approximate pragmatic use of language through integrating various skills and components and utilizing sociolinguistic rules involved in actual communication (Oller, 1983; Spolsky, 1975; Farhady, 1980, among others). Dictation and cloze are two well-known types of IN tests. Contrary to their merits, their reliability and validity were questioned by Alderson (1978) in that they do not measure what they are supposed to measure, and by Farhady (1994, 1995) in that they violate the independence of items which

(DP), integrative (IN), pragmatic (P), and communicative (C) tests are the outcomes of these theoretical perspectives. However, following Vollmer and Sang (1983), Bachman and Palmer (1996) and Kunnan (1999) state that the nature of language proficiency on the whole is not yet at all clear and the universe of language proficiency have not been fully explored. The purpose of this paper, then, is to investigate the effect of two of the factors influencing language performance and probably its underlying structure. More specifically, this paper aims at exploring the effect of the level of language proficiency, test method, and their possible interaction on the underlying structure of language proficiency.

Overview

Teaching and testing are theoretically and practically interrelated. Theoretically, they have been influenced by the principles of linguistics, psychology, and other language related fields. Practically, they are complementary because both aim at optimizing the efficiency of education (Eble, 1972; Gronlund; 1976, Farhady, 1983). Regarding language testing in the last three decades, Spolsky's 1978 classic categorization of language testing eras, i.e., pre-scientific, strucrural-psychometric, and integrative-sociolinguistic has been the dominant classification, and only one era which Bachman (1990) calls it psychomertic-communicative is added to language testing advancements.

The Pre-scientific period was identified with subjective judgment where testing was considered an art rather than a science (Heaton, 1988). Main

ability.

Introduction

Human being, as an intellectual creature, possesses many intricate capabilities the most complex of which is language. This capacity in general, and its underlying structure in particular, has been a polemic issue in research studies for over a century (Spearman, 1904). Research has resulted in various theoretical and methodological developments. Theoretically, different models are suggested ranging along a continuum with multidimensional models at the one end, unidimensional one at the other, and some moderate ones in between. (Carroll, 1961; Spolsky, 1973, 1995; Oller, 1983; Vollmer, 1983; Skehan, 1988; McNamara, 1995; Farhady, 1997, to name a few). A number of terms and concepts, even sometimes contradictory, such as divisible, indivisible hypotheses (each with weak and strong versions), partially divisible, mutually exclusive, two complementary hypotheses, (Oller & Hinofotis, 1980; Davies, 1991), and hierarchical skills theory (Sang et al., 1986), to name some, have appearad in the literature. Futher, recent developments have revealed language ability is not independent of the factors outside the scope of language itself. Test taker characteristics, test rubrics, test method, item format, and the level of language proficiency are some of the factors contributing to the underlying structure of language ability (Vollmer, 1983; Hughes & Porter, 1983; Alderson, 1986; Milanovic, 1988; Anivan, 1991).

Methodologically, various techniques have been developed to construct different tests following the existing theoretical models. Discrete-point

فصلنامه علمی - پژوهشی علوم انسانی دانشگاه الزهرا(س)

سال نهم، شماره ۲۹، بهار ۱۳۷۸

Test Method, Level of Language Proficiency, and Underlying Structure of Language Ability

Dr. Hossein Farhady*

Gholam-Reza Abbasian**

Abstract

This study investigated the underlying structure of language ability in relation to the levels of language proficiency and test method. 423 undergraduate English majors were divided into three groups of elementary, intermediate, and advanced levels through administering TOEFL. Then, a number of tests corresponding to the different levels of proficiency were developed and administered.

The results revealed an ascending order in the magnitude of correlation coefficients from the elementry to the advanced levels of language proficiency. Furthermore, factor analysis resulted in three distinct factors for the elementary and intermediate levels but only two factors for the advanced level. It was concluded that language proficiency might have a level-dependent rather then a static nature. Based on the findings, it was proposed that teaching and testing language might be more appropriate if the learners are treated differently on the basis of their level of language

* Iran University of Science and Technology

** Tarbiat Modarres University